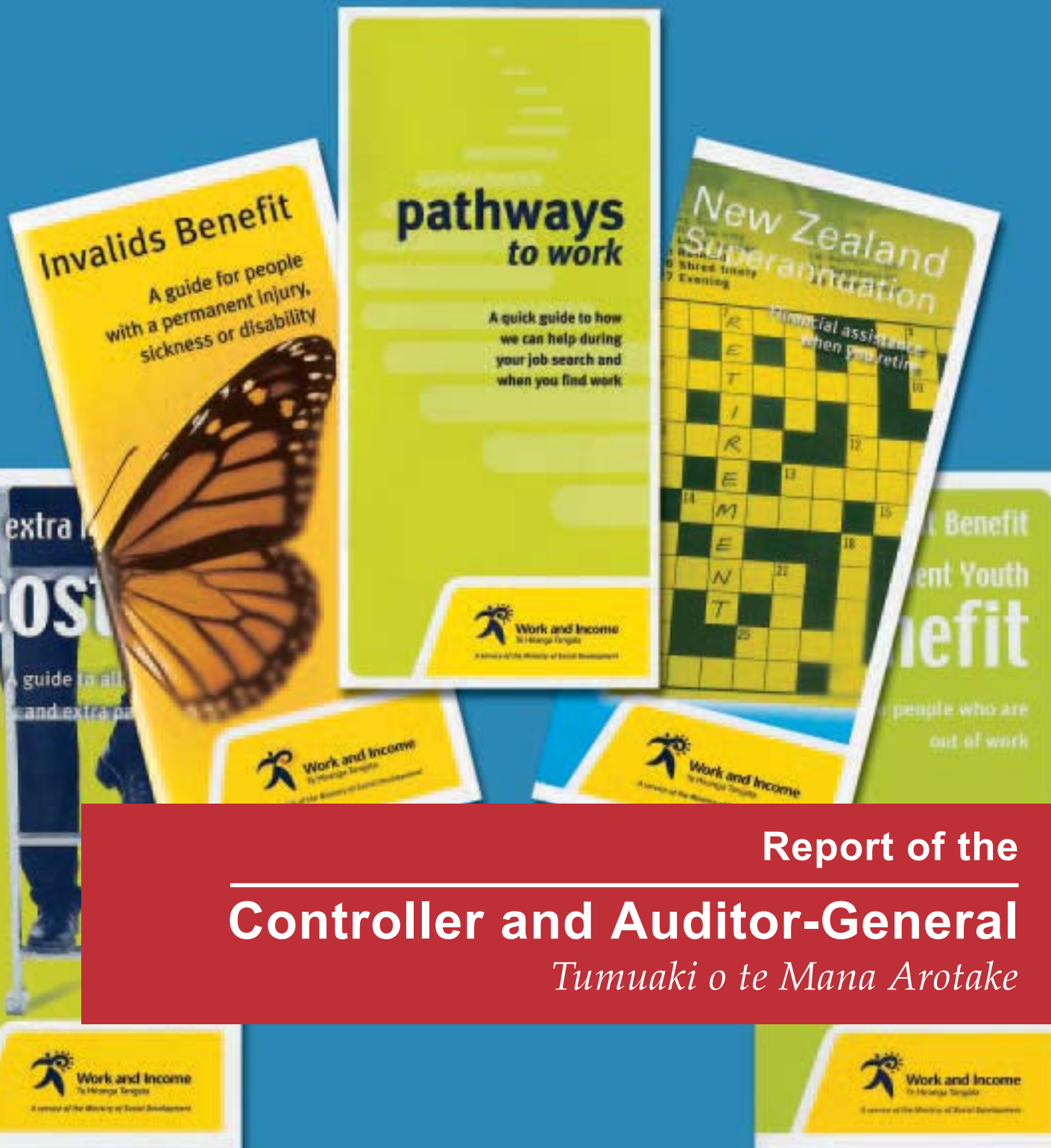


# Social Security Benefits: Accuracy of Benefit Administration



Report of the  
**Controller and Auditor-General**  
*Tumuaki o te Mana Arotake*



**Report of the**

**Controller and**

**Auditor-General**

*Tumuaki o te Mana Arotake*

**Social Security Benefits:**

**Accuracy of**

**Benefit Administration**

**December 2003**

**Note:**

*There is a distinction between “benefits” (which are mostly “monetary benefits” payable under the Social Security Act 1964) and “pensions” (principally New Zealand Superannuation payable under the New Zealand Superannuation Act 2001 – which accounts for nearly half of all income support payments). We have treated both as “benefits” because the Ministry of Social Development employs the same systems and methods to administer them.*

**ISBN 0-478-18104-3**

## Foreword

One of my key objectives for my term as Auditor-General is to promote improvements in the quality of performance reporting in the public sector. I will be building on work begun by my predecessor who published a key guidance document entitled *Reporting Public Sector Performance*. I am especially keen to see the principles contained in that document reflected in substantial improvements in the quality of many current output performance measures used by government departments and other public entities.

My objective has acquired even greater importance as a result of the initiative known as *Managing for Outcomes*, which was approved by the Cabinet in December 2001. This initiative requires departments to adopt a strategic and outcome-focused approach to planning, managing, and reporting, while continuing to focus on delivering outputs. As a consequence, output performance measures now need to form part of a comprehensive and coherent information set that describes each department's total performance in the context of its strategic environment.

In 2002-03, the Ministry of Social Development paid \$11,743 million to over 800,000 beneficiaries. Given the very large sums involved, it is clearly important to ensure that payments are made accurately. An avoidable error rate of only 1% would represent a sum of approximately \$120 million.

Our audit identified a number of areas in which we believe the information on benefit accuracy currently collected and published by the Ministry could be extended and improved.

I thank the Ministry's staff for their co-operation in the conduct of the audit. I also extend my particular thanks to Professor Stephen Haslett of Massey University, who acted as our advisor on a number of statistical issues.



K B Brady  
Controller and Auditor-General

1 December 2003





## CONTENTS

### Contents

|  | <i>Page</i> |
|--|-------------|
| <b>Conclusions and Recommendations</b>                                 | <b>6</b>    |
| <b>Part One: Nature and Scope of Our Audit</b>                         | <b>15</b>   |
| Why We Carried Out Our Audit   | 17          |
| Our Objectives   | 19          |
| Our Mandate for the Audit  | 19          |
| What We Did  | 20          |
| What We Did Not Look At  | 20          |
| <b>Part Two: Organisation and Management of Benefit Administration</b> | <b>23</b>   |
| The Ministry of Social Development                                     | 25          |
| Types of Benefit   | 27          |
| The Benefit System   | 29          |
| The Work of a Case Manager   | 30          |
| <b>Part Three: Systems and Processes for Ensuring Benefit Accuracy</b> | <b>33</b>   |
| Introduction   | 35          |
| The Benefit Processing System (SWIFTT)                                 | 35          |
| Applying for a Benefit and Supplying Information                       | 36          |
| <b>Part Four: The Pivotal Role of the Case Manager</b>                 | <b>41</b>   |
| Introduction   | 43          |
| Case Managers' Caseloads   | 43          |
| "5+5 Checks" on Case Managers' Performance                             | 48          |
| Role of Team Coaches   | 51          |
| Case Manager Performance Assessment                                    | 52          |



|  | <i>Page</i> |
|--|-------------|
| <b>Part Five: Accuracy Measurement and Reporting</b>           | <b>55</b>   |
| Defining Accuracy  | 57          |
| The Ministry's Performance Measure for Accuracy                | 57          |
| The Accuracy Reporting Programme –                             |             |
| What It Measures and What It Doesn't                           | 58          |
| Unverifiable Items   | 61          |
| Uncertainty of ARP Results                                     | 63          |
| Measuring and Reporting the Size of Errors and Fraud           | 68          |
| Identifying and Reporting the Causes of, and                   |             |
| Hardship Caused by, Errors                                     | 70          |
| <b>Part Six: Other Opportunities for Improvement</b>           | <b>71</b>   |
| Introduction   | 73          |
| Consequences of the Ministry's Structure                       | 73          |
| Obtaining More Complete Information on the                     |             |
| Population of Potential Beneficiaries                          | 76          |
| <b>Figures</b>   |             |
| 1 Expenditure on Income Support 2002-03                        | 17          |
| 2 The Ministry's Geographic Structure                          | 26          |
| 3 The Ministry's Functional Structure                          | 27          |
| 4 The Three Tiers of Benefit                                   | 28          |
| 5 Components of the Benefit System                             | 29          |
| 6 Activities That Can Affect Case Manager Accuracy             | 31          |
| 7 Applications for Additional Assistance 1999 to 2002          |             |
| as a Proportion of Working-age Beneficiaries                   | 45          |
| 8 Options for Managing Caseloads – To Specialise or Not?       | 46          |
| 9 Benefit Accuracy Information – 1996 National Survey of       |             |
| Eligibility for the Accommodation Supplement                   | 60          |
| 10 Regional Accuracy Results for 2001-02 Against the           |             |
| National Target as Measured by ARP Excluding Unverifiables     | 62          |
| 11 Ranges of Regional Accuracy Results for 2001-02 Against the |             |
| National Target  | 64          |
| 12 Funding of Service Delivery and Specialist Services         | 74          |
| 13 Ministry Activities to Improve "Uptake" of Benefits         | 77          |



# Conclusions and Recommendations

In 2002-03, the Ministry of Social Development (the Ministry) was responsible for the administration of nearly \$12,000 million of Crown funding for income support for beneficiaries. The Ministry has systems to measure payment accuracy, and the measures are audited and reported annually.

The Ministry's obligation is to pay benefits correctly on the basis of the information available to it (whether the information was provided by or obtained from the applicant/beneficiary, or was otherwise known to it). It measures accuracy in terms of this obligation.

We think this measure of accuracy is of only limited use in enabling Parliament to assess whether or not the Ministry is doing a good job. The Ministry does not currently collect information in other areas that we think are relevant to the general issue of accuracy. For example, it does not seek to estimate the total cost of payment errors, whether in terms of the amounts of benefit overpaid or underpaid, or in terms of the true cost of "fixing" the errors.

Because of the limitations of the information currently available, we have been unable to form a clear view about whether or not the Ministry is performing well. In saying this, we acknowledge that the Ministry has been meeting the performance targets in its recent *Departmental Forecast Reports*. We also acknowledge that Ministry staff have competing priorities, and that accuracy is only one dimension of their performance. Nonetheless, given the very large sums of money involved, we are firm in our view that more and better information is required.

## Our Audit and Recommendations

In view of the very substantial sums involved and the implications of inaccurate benefit payments for the welfare of both taxpayers and beneficiaries, we considered it necessary to supplement our annual audit with a special audit of the soundness of the Ministry's systems for benefit administration and accuracy measurement.





## CONCLUSIONS AND RECOMMENDATIONS

We have made 14 recommendations in this report to highlight areas where improvements could be made.

### The Benefit Processing System

The Ministry's computer system, SWIFTT, while old, is sound in terms of accurately processing benefits. It has some constraints that limit its use to identify accuracy. The Ministry has developed some new functions, and there is scope for the system to be enhanced further.

#### **Recommendation 1** (page 36)

We recommend that –

The Ministry continues to explore further enhancements to SWIFTT involving automatic checks, alerts to case managers, compulsory fields to provide explanations of decisions, and the production of management information.

### The Extent and Causes of Under- and Over-payments

Under- or over-payments that arise outside of the Ministry's obligation (for example, because the applicant provides incorrect information) can still have consequences for the Ministry or the beneficiary, or both. There could be benefits in learning more about why they happen, with a view to devising ways to avoid them, or to identify them as soon as possible.

#### **Recommendation 2** (page 40)

We recommend that –

The Ministry continues to explore cost-effective ways of extracting and analysing data about the extent and causes of under- and over-payments, and uses the results of the analysis to devise ways to avoid under- and over-payments or to identify them as soon as possible.





## CONCLUSIONS AND RECOMMENDATIONS

### Case Managers' Caseloads

The Ministry's arrangements for allocating resources and devolving responsibility for managing them are consistent with giving managers appropriate responsibility for what they control and making them accountable for results. However, benefit administration is likely to be improved if reliable assessments of the benefits and risks to accuracy of different approaches to improving accuracy were routinely shared with other regions, rather than remaining as lessons held within particular regions.

Because accuracy results are only reported annually to the national level, any region that adopts an approach that turns out to alter accuracy will not come to national management's attention in a timely way. In the meantime, if the impact is to reduce accuracy, beneficiaries may be disadvantaged for a considerable length of time.

Regions are resourced according to the number of clients. Those regions with a greater than average number of beneficiaries with complex circumstances could be disadvantaged, because the additional work arising from dealing with these beneficiaries could leave them less able than other regions to meet their accuracy target.

#### **Recommendation 3** (page 48)

We recommend that –

The Ministry continues to promote sharing of information among regions and with National Office on approaches to staff deployment and managing caseloads.

#### **Recommendation 4** (page 48)

We recommend that –

For those regions that have a high number of beneficiaries with complex circumstances, the Ministry explores whether a relatively higher allocation to reflect the workload associated with complexity is justified.





## CONCLUSIONS AND RECOMMENDATIONS

### Checks on Case Managers' Performance

Key Performance Indicators (KPIs) are a central feature of the Ministry's culture. Case managers have clear, measurable targets to aim for in their day-to-day work.

The checks undertaken at service centres (known as 5+5 checks) provide valuable information on accuracy, and are an effective means of achieving and maintaining accuracy. In those regions that look at the results across service centres, they also provide useful comparative information.

A lack of more high-level comparative analysis of the results – particularly between service centres dealing with similar beneficiary profiles and problems – is a missed opportunity to compare performance and to transfer lessons.

#### **Recommendation 5** (page 50)

We recommend that –

The Ministry investigates the possibility of using aggregated 5+5 data at regional level to draw inferences about regional performance through the use of appropriate statistical techniques such as double sampling.

### The Use of Team Coaches

Team coaches provide important and responsive support and training to the case managers. We consider that not sharing information on how regions use team coaches is (as with the 5+5 checks) a missed opportunity to compare performance and to transfer lessons.

#### **Recommendation 6** (page 52)

We recommend that –

The Ministry continues to promote sharing of information among regions and with National Office on approaches to the use of team coaches.





## CONCLUSIONS AND RECOMMENDATIONS

### Assessing Case Managers' Performance

The Ministry has a systematic approach to linking individual performance of its case managers to its KPI for processing accuracy. However, the way some regions assess performance creates a risk that staff will not see accuracy as a high priority, potentially leading to an inconsistent level of accuracy between case managers and regions.

#### **Recommendation 7** (page 53)

We recommend that –

The Ministry requires all regions to assess the performance of staff consistently. The chosen method should provide an incentive to staff to accord accuracy an appropriately high priority.

### Measuring and Reporting Accuracy

The Accuracy Reporting Programme (ARP) measures and reports accuracy on the basis of the Ministry's statutory obligations and Purchase Agreement. It does not, and is not intended to, provide a wider measure of benefits being paid accurately to all who are eligible to apply – which would entail a separate special exercise in data collection and analysis.

Some of the cases selected as part of the ARP sample cannot be assessed as correct or incorrect because files or parts of files are found to be missing. At present the Ministry excludes these cases – known as “unverifiabiles” – from the sample. This practice could lead to an incorrect estimate of the overall level of accuracy, because cases for which papers cannot be found may well have an error rate that is different from the rate for cases that are well documented.

By changing this practice to include unverifiable cases as errors, there would be a strong incentive for the Ministry to reduce the unverifiabiles to as low a level as possible. However, there would also be a risk that the results could under-estimate the overall level of accuracy, since it is unlikely that all unverifiable items contain errors.

An alternative approach would be for the Ministry to clearly explain unverifiabiles and disclose their level in its annual report.





## CONCLUSIONS AND RECOMMENDATIONS

### **Recommendation 8** (page 63)

We recommend that –

The Ministry explains unverifiables and discloses their level in its annual report. The precise form of the disclosure should be agreed with the Audit Office as part of the annual audit.

## Uncertainty of Measurement Results

The ARP was not designed as a statistically valid instrument for measuring regional performance. Consequently, because the ARP estimates the level of accuracy of the total “population” of benefits on the basis of the accuracy of a relatively small random sample of those benefits, any use of ARP results in respect of regions needs to be approached with caution. Once down to regional level, the samples are small, and the smaller the sample on which an estimate is based, the greater the level of uncertainty of the estimate.

In our view, the relative levels of accuracy between different parts of the country are important. Regions’ ARP results provide an imperfect measure of accuracy, but, in the absence of something better, it is acceptable for regional managers to use them as one of their operational tools – so long as they are clear about the uncertainties of the data they are using.

### **Recommendation 9** (page 68)

We recommend that –

The Ministry continues to give the regions their ARP results, but in a form similar to Figure 11 on page 64 – showing each region’s data at a 95% confidence level, and comparative data of other regions.

### **Recommendation 10** (page 68)

We recommend that –

The Ministry provides all Regional Commissioners and Regional Operations Managers with training on the nature of sampling error and the appropriate interpretation of statistical estimates that include confidence intervals.





## CONCLUSIONS AND RECOMMENDATIONS

### Identifying the Size of Errors and Fraud

The ARP does not estimate the amount of under- or over-payments. Other available information and our own analysis suggest that the risk of large errors is low. However, it is important to undertake a specific exercise periodically to estimate the amount of over-payments (including fraud). The estimate would also provide the Ministry with a factual basis on which to estimate whether the current level of expenditure on Benefit Control (\$38.3 million in 2003-04) yields the greatest cost/benefit.

#### **Recommendation 11** (page 70)

We recommend that –

The Ministry regularly performs a risk-sizing exercise to estimate the amount of over-payments.

### The Causes of, and Hardship Caused by, Errors

A more systematic collection of information on errors, their size and cause, and how they were found would provide useful indicators to help management assess for policy purposes:

- priorities for focusing effort to achieve better accuracy – by identifying processing errors early and by avoiding them; and
- the hardship that such errors might be causing beneficiaries.

#### **Recommendation 12** (page 70)

We recommend that –

The Ministry continues to explore the collection and analysis of information on errors, their size and cause, and how they were found, and to link this work with enhancements to its information technology systems.





## CONCLUSIONS AND RECOMMENDATIONS

### Consequences of the Ministry's Structure

The Ministry views the Service Delivery and Specialist Services components of its benefit administration separately. Viewing them in this way runs the risk that potential cost/benefit opportunities from examining the various components together will not be identified.

#### **Recommendation 13** (page 76)

We recommend that –

The Ministry treats all its processes for administering benefits as components of an integrated system. It should periodically re-estimate the optimum balance of effort between the different components in order to achieve the most cost-beneficial outcome. The method of estimation should consider all costs and benefits, including those incurred by or realised by beneficiaries, rather than being confined to those of the Ministry.

### Potential Beneficiaries

In our view, the Ministry requires fuller information on the potential population of beneficiaries. Its activities to identify people who are eligible for a benefit but have not applied for it are likely to be more effective if the characteristics of the target population are known.

#### **Recommendation 14** (page 78)

We recommend that –

The Ministry periodically undertakes exercises to estimate the number of people who are potentially eligible for social security assistance but who have not applied. For a supplementary benefit such as the Accommodation Supplement, the exercise would be likely to require information on personal income that might best be obtained from the Inland Revenue Department. The Ministry should investigate the feasibility of undertaking data matching or data extraction exercises that would yield the necessary information.





## Part One

# Nature and Scope of Our Audit

### Sickness Benefit

For people who are temporarily off work because of sickness, injury, pregnancy or disability



Work and Income  
Te Hiraonga Tangata

A service of the Ministry of Social Development

### Relationships and income support

Things you need to know about relationships when getting income support



Work and Income  
Te Hiraonga Tangata

A service of the Ministry of Social Development

### Orphans and Unsupported Childs Benefits

A guide for caregivers who are raising someone else's child



Work and Income  
Te Hiraonga Tangata

A service of the Ministry of Social Development





## Why We Carried Out Our Audit

- 1.1 The Ministry of Social Development (the Ministry) has systems to measure payment accuracy and the measures are audited and reported annually. (See Part 5 on pages 55-70 for comment on the specific measures.)
- 1.2 In view of the very substantial sums involved and the implications of inaccurate benefit payments for the welfare of both taxpayers and beneficiaries, we considered it necessary to supplement our annual audit with a special audit of the soundness of the Ministry's systems for benefit administration and accuracy measurement.

### Substantial Sums of Public Money Involved

- 1.3 In 2002-03, the Ministry was responsible for the administration of nearly \$12,000 million of Crown funding for income support for beneficiaries (see Figure 1).

*Figure 1  
Expenditure on Income Support 2002-03*

| Benefit*                   | Expenditure<br>\$million |
|----------------------------|--------------------------|
| New Zealand Superannuation | 5,642                    |
| Domestic Purposes          | 1,520                    |
| Unemployment               | 1,274                    |
| Invalids                   | 914                      |
| Accommodation Supplement   | 706                      |
| Sickness                   | 421                      |
| Student Allowances         | 387                      |
| Other                      | 879                      |
| <b>Total</b>               | <b>11,743</b>            |

\* Includes supplementary benefits paid with the main benefits listed.

Source: *Ministry of Social Development Annual Report 2002-2003*, parliamentary paper G.60, 2003, pages 131-132.



## NATURE AND SCOPE OF OUR AUDIT

### *Implications for Taxpayers and Beneficiaries*

- 1.4 The Ministry's *Annual Report 2002-2003* records that, across the whole year, it provided assistance and support to:
- almost 840,000 people who received an income support benefit or pension;
  - almost 215,000 people who received a Student Loan or Student Allowance; and
  - 1.35 million people who received a Community Services Card.
- 1.5 Income support is intended to enable people to participate in their communities and the labour market, and therefore has a vital role in meeting a number of the Government's key goals – including:
- growing an inclusive, innovative economy for the benefit of all;
  - providing strong social services;
  - improving skills; and
  - reducing inequalities in health, education, employment and housing.<sup>1</sup>
- 1.6 Inaccurate benefit payments can result in costs to both the Ministry and beneficiaries.
- 1.7 The Ministry incurs direct costs from the administrative actions associated with identifying and correcting any errors. These actions can include:
- reviewing and correcting the amount of benefit payable;
  - establishing any over-payment as a debt owed to the Crown;
  - processing an application by the beneficiary for a review of (or an appeal against) the decision to establish the over-payment; and/or
  - attempting to recover the debt.
- 1.8 Opportunity costs arise from Ministry staff having to spend time and effort on correcting benefits instead of undertaking other work (such as employment promotion) that might also assist a beneficiary.

<sup>1</sup> Ministry of Social Development, *Departmental Forecast Report 2002-2003*, parliamentary paper E12.FR2002, page 5. [www.msd.govt.nz/documents/publications/msd/dfr2002.pdf](http://www.msd.govt.nz/documents/publications/msd/dfr2002.pdf)



- 1.9 Beneficiaries can suffer hardship when payments are made incorrectly. Under-payments deprive them of income to which they are entitled. Over-payments generally must be repaid, placing financial pressure on beneficiaries – often for extended periods.
- 1.10 It is also important that Parliament is assured that taxpayer-funded benefits are being paid to the people whom Parliament intended and in the intended amounts.

### Our Objectives

- 1.11 In conducting our audit, we sought to provide Parliament with assurance that:
- the Ministry has effective systems and methods for ensuring benefit accuracy; and
  - Parliament can place reliance on the performance data that the Ministry reports.

### Our Mandate for the Audit

- 1.12 We conducted the audit under section 16(1)(a) of the Public Audit Act 2001, which provides that –
- The Auditor-General may at any time examine –*
- (a) the extent to which a public entity is carrying out its activities effectively and efficiently.*
- 1.13 When conducting the audit, we have at all times kept in mind the need to observe section 16(4) of the Public Audit Act. That is, we must limit our examination under section 16(1)(a) to the extent to which the Ministry is carrying out its activities effectively and efficiently in a manner consistent with applicable Government policy to which the Ministry is required to adhere.
- 1.14 We were also conscious that the Ministry's objectives to assess benefit eligibility are determined by statute (see paragraphs 2.8-2.10 on page 29).



## NATURE AND SCOPE OF OUR AUDIT

### What We Did

- 1.15 We interviewed the Ministry's National Office staff and staff in five regions (Auckland South, East Coast, Nelson, Waikato, and Wellington). Staff in the regions included:
- Regional Commissioners;
  - Regional Operations Managers;
  - service centre managers;
  - team coaches; and
  - case managers.
- 1.16 We reviewed the type of reporting carried out for both internal and external use. We also visited a training centre for case managers.

### What We Did Not Look At

- 1.17 If an inaccuracy results in an over-payment of a benefit, the beneficiary will generally need to repay the over-paid amount. The Ministry classifies over-payments into four types:
- fraud and abuse;
  - data matching;
  - over-payments/other; and
  - recoverable special needs grants and advances.
- The Ministry reports debt under either “fraud” or “data match debt” categories.
- 1.18 In 2001 the Ministry commissioned a research economist, Dr Anne Heynes, to undertake a study on the causes of innocent over-payment debt. Her report<sup>2</sup> identified, among other items, areas requiring further research – including:
- the circumstances or reasons that led to the establishment of innocent over-payment debt (noting that survey methods might be required as this information cannot be accurately determined from available data); and

2 Heynes, A, *The Causes of Innocent Overpayment Debt: A research report prepared for the Ministry of Social Development*, Ministry of Social Development, October 2001. <http://www.msd.govt.nz/publications/sector-policy/innocent-overpayment-debt.pdf>



- specifically, whether each new such over-payment that arose through beneficiary or staff error requires new information to be collected – this information would also need to be obtained from a survey.
- 1.19 We have not sought to explore debt management in our report, but note that the research requirements that Dr Heynes identified mirror some of our conclusions and recommendations.
- 1.20 Our audit also did not look at certain well-publicised issues (briefly outlined below) of:
- entitlement to third-tier benefits; and
  - application of Court decisions.

### Entitlement to Third-tier Benefits

- 1.21 The Ministry and advocacy groups have conflicting views about entitlement to certain third-tier benefits – particularly the Special Benefit. We have not addressed this issue because the Ministry is currently looking into variability between its service centres when granting third-tier benefits, in the light of submissions by advocacy groups.

### Application of Court Decisions

- 1.22 The law relating to social security benefits is complex. The statutory basis is primarily the Social Security Act 1964, on which there is a continuously developing body of case law. The Act is also subject to frequent amendment.
- 1.23 One widely publicised case concerned the 1996 decision of the Court of Appeal in the case *Ruka v Department of Social Welfare*.<sup>3</sup> The decision substantially altered the test to be used to determine whether a beneficiary was in “a relationship in the nature of a marriage”. It had important social security and legal consequences for a person considered to be in such a relationship.

3 [1997] 1 NZLR 154.



## NATURE AND SCOPE OF OUR AUDIT

- 1.24 In September 2000 the (then) Associate Minister for Social Services commissioned a barrister, Frances Joychild, to undertake a review in response to public criticism of the way in which the former Department of Social Welfare and then the Department of Work and Income had responded to the decision. Ms Joychild's report<sup>4</sup> found that:
- investigators had not changed their decision-making behaviour to match the interpretation outlined in the Court of Appeal decision; but
  - since December 2000 the decision had been correctly applied.
- 1.25 One contributory reason had been the Departments' failure to amend manuals and guidance materials between November 1996 and mid-2000 to reflect the decision. Ms Joychild's report highlighted problems in ensuring consistency of interpretation between service centres.
- 1.26 We did not examine this matter because the Ministry was concurrently responding to Ms Joychild's report.<sup>5</sup>

4 Joychild, F, *Report to the Minister for Social Services, Review of Department of Work and Income Implementation of the Court of Appeal Decision, Ruka v Department of Social Welfare* [1997] 1 NZLR 154, Ministry of Social Development, June 2001.

5 *Ministry of Social Development Response to the Joychild Report* [www.msd.govt.nz/publications/response\\_joychild\\_report/](http://www.msd.govt.nz/publications/response_joychild_report/) or [www.msd.govt.nz/publications/index.html](http://www.msd.govt.nz/publications/index.html) *Implementation of the Ruka Decision*, Report to the Minister of Social Services and Employment, 2002.



## Part Two

# Organisation and Management of Benefit Administration





### The Ministry of Social Development

- 2.1 The Ministry is a combination of two previous organisations – the Department of Work and Income (otherwise known as Work and Income New Zealand or WINZ) and the Ministry of Social Policy.
- 2.2 WINZ was originally created in 1998 through a merger of the Employment Service division of the Department of Labour and the Income Support business unit of the Department of Social Welfare. WINZ was given the task of helping beneficiaries to achieve stable employment and self-sufficiency, as well as administering benefit payments. The Ministry has now taken over both of these roles.
- 2.3 Processing benefit applications and payments is an important part of the Ministry's role and forms the majority of work conducted at its service centres. However, the Purchase Agreement between the Minister of Social Services and Employment and the Chief Executive of the Ministry puts greater emphasis on work placement – reflecting a priority on placing beneficiaries into stable and sustainable employment. The current Key Performance Indicators (KPIs) for service delivery include two relating to the processing of benefits – one for accuracy and one for timeliness – compared with a larger number of KPIs that are directed at work placement.

#### Geographic Structure

- 2.4 Figure 2 on the next page illustrates the manner in which the Ministry's offices are structured throughout the country.

#### Service Centre Staffing

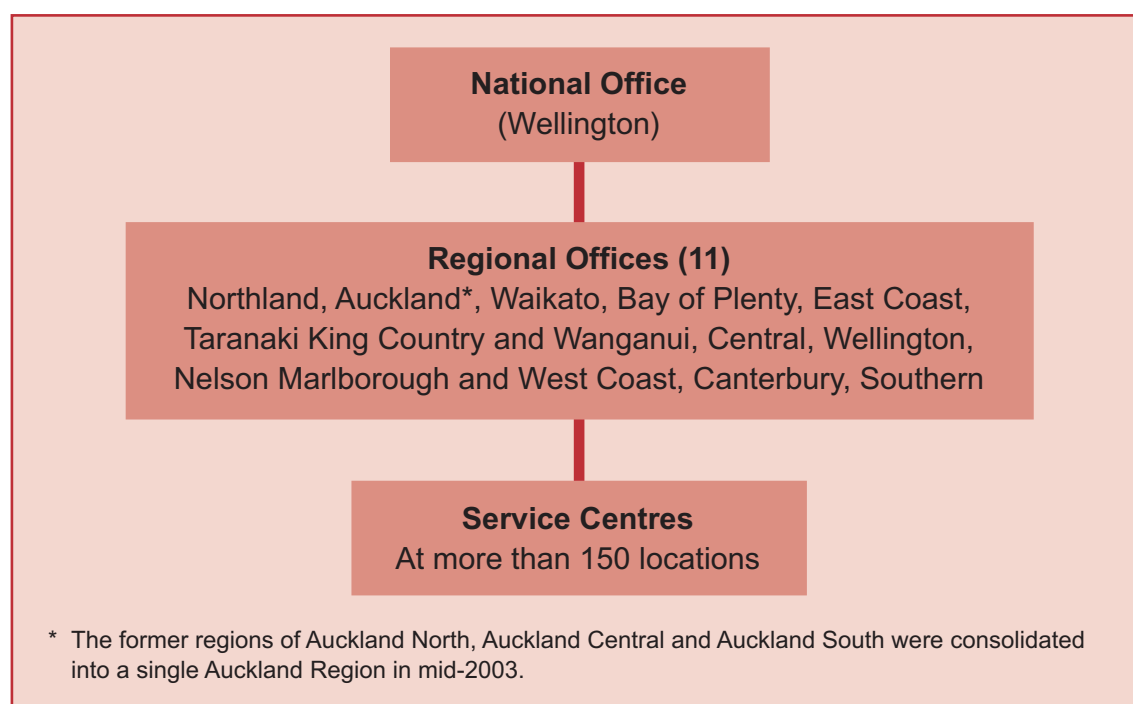
- 2.5 Within each service centre, the essential levels of operational staff are:
- **Service Centre Manager** – who leads the service centre team, and is responsible for the quality of service delivery and the achievement of business plan objectives.
  - **Team Coaches** – who provide coaching, feedback, tuition and support to case managers to enable them to grow in their role and achieve a high level of service and excellent outcomes for their clients.



## ORGANISATION AND MANAGEMENT OF BENEFIT ADMINISTRATION

- **Case Managers** – who manage a caseload of clients, providing ongoing case management services to individual Work and Income clients within their caseload.

*Figure 2*  
*The Ministry's Geographic Structure*



### Functional Structure

**2.6** The Ministry is organised into two key operational areas (see Figure 3 on the opposite page):

- *Service Delivery (Work and Income)* is done by case managers who, in respect of administering benefit payments, are primarily responsible for ensuring that what the Ministry pays a beneficiary is accurate from the outset. Case managers are responsible for assisting beneficiaries to achieve stable employment.
- *Specialist Services* provides a range of services to discrete population groups – Studylink, the Community Services Card, International Services, War Pension Services, Benefit Control (including the National Data Match Centre) and Debt. Area Benefit Control teams have responsibility for detecting over-payments resulting from fraud. They also have intervention strategies to ensure that the client is in receipt

## ORGANISATION AND MANAGEMENT OF BENEFIT ADMINISTRATION

of the correct benefit. The National Data Match Centre matches information with other government agencies and establishes over-payments, where it is clear the client was not entitled to a benefit or should have been on a lower rate of benefit.

**Figure 3**  
*The Ministry's Functional Structure*



## Types of Benefit

**2.7** Benefit types are structured in three tiers (see Figure 4 on the next page):

- First-tier benefits comprise the core entitlement based on the beneficiary's reason for the benefit application – for example, Domestic Purposes Benefit, Unemployment Benefit, Sickness Benefit, and Invalids Benefit.
- Second-tier benefits – known as “Supplementary Benefits” – are an entitlement where the beneficiary has a particular set of circumstances that incur extra costs. These benefits include the Accommodation Supplement, Disability Allowance and Child Disability Allowance.
- Third-tier benefits provide assistance for people who have a (generally) temporary need, and are an acknowledgement that the core and supplementary benefits will not always enable the beneficiary to meet costs that result from an emergency or unexpected situation. Payment of a third-tier benefit is discretionary and is not an entitlement until such time as a decision to exercise the discretion has been taken. Third-tier benefits include recoverable and non-recoverable Special Needs Grant, Advances, and Special Benefit.

## ORGANISATION AND MANAGEMENT OF BENEFIT ADMINISTRATION



Figure 4  
*The Three Tiers of Benefit*

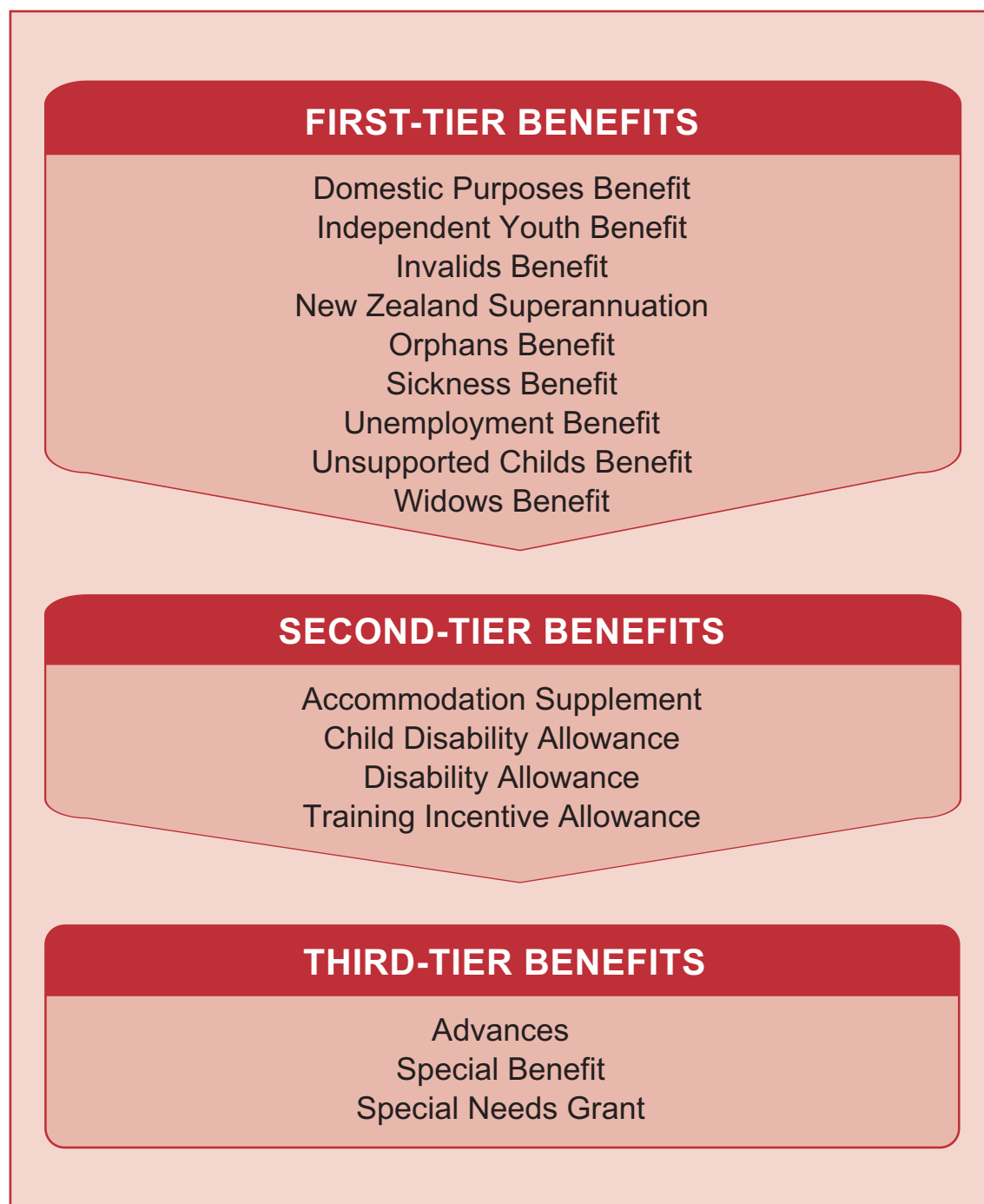
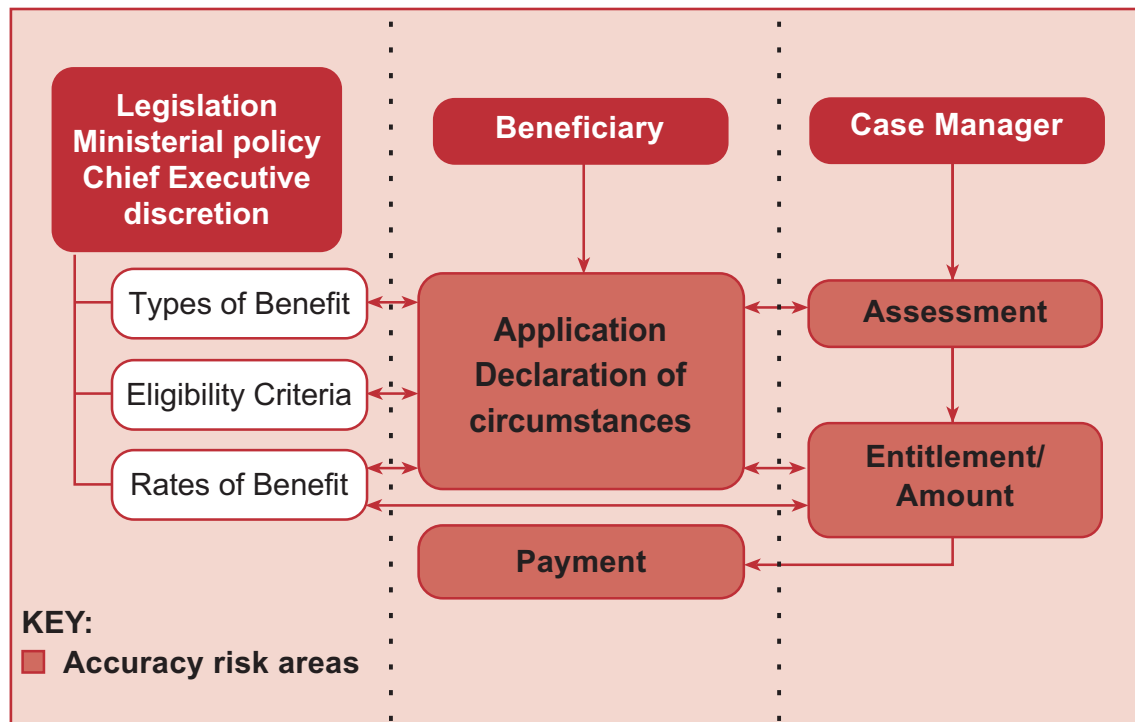




Figure 5  
Components of the Benefit System



## The Benefit System

- 2.8 The provisions of the Social Security Act 1964 broadly dictate the way in which the benefit system works. Benefit types, eligibility criteria, and rates are set by (one or more of) legislation, Ministerial policy, and the exercise of discretion by the Ministry's Chief Executive.
- 2.9 Under section 12(1) of the Social Security Act, the Ministry can grant benefits only to people who make a "claim" for a benefit. However, when making application, the claimant is not required to specify a particular monetary benefit to be eligible for it.<sup>6</sup>
- 2.10 The Ministry is not legally obliged to seek out people who may be eligible for a benefit but have not applied. However, every person eligible to receive a benefit and who applies for it should expect to be paid the amount they are entitled to receive. After receipt of the application, the Ministry's Chief Executive has a legal duty to "investigate"<sup>7</sup> the claim to a benefit. It is only after this investigation that an "entitlement" to a benefit is established.

<sup>6</sup> *Chief Executive of the Department of Work and Income v Scoble* [2001] NZAR 1011.

<sup>7</sup> The Ministry prefers to use the term "assess". We use that term in our report.

## ORGANISATION AND MANAGEMENT OF BENEFIT ADMINISTRATION



- 2.11 Consequently, application, assessment and entitlement comprise the key components of the benefit system – as illustrated in Figure 5 on the previous page, which indicates the key areas where risks to accuracy arise.

### The Work of a Case Manager

- 2.12 Case managers make most of the decisions on payments. The main responsibility for ensuring that payments are accurate therefore lies with case managers.
- 2.13 Figure 6 on the opposite page illustrates eight key Service Delivery activities and their various components that can have an effect on the work of the case manager. For example, case manager training is carried out at three levels:
- the service centre – involving the case manager and their team coach;
  - regional level – where the regional trainer and regional operations manager plan the training of service centre staff in their region; and
  - national level – which deals with induction training and training in new initiatives.

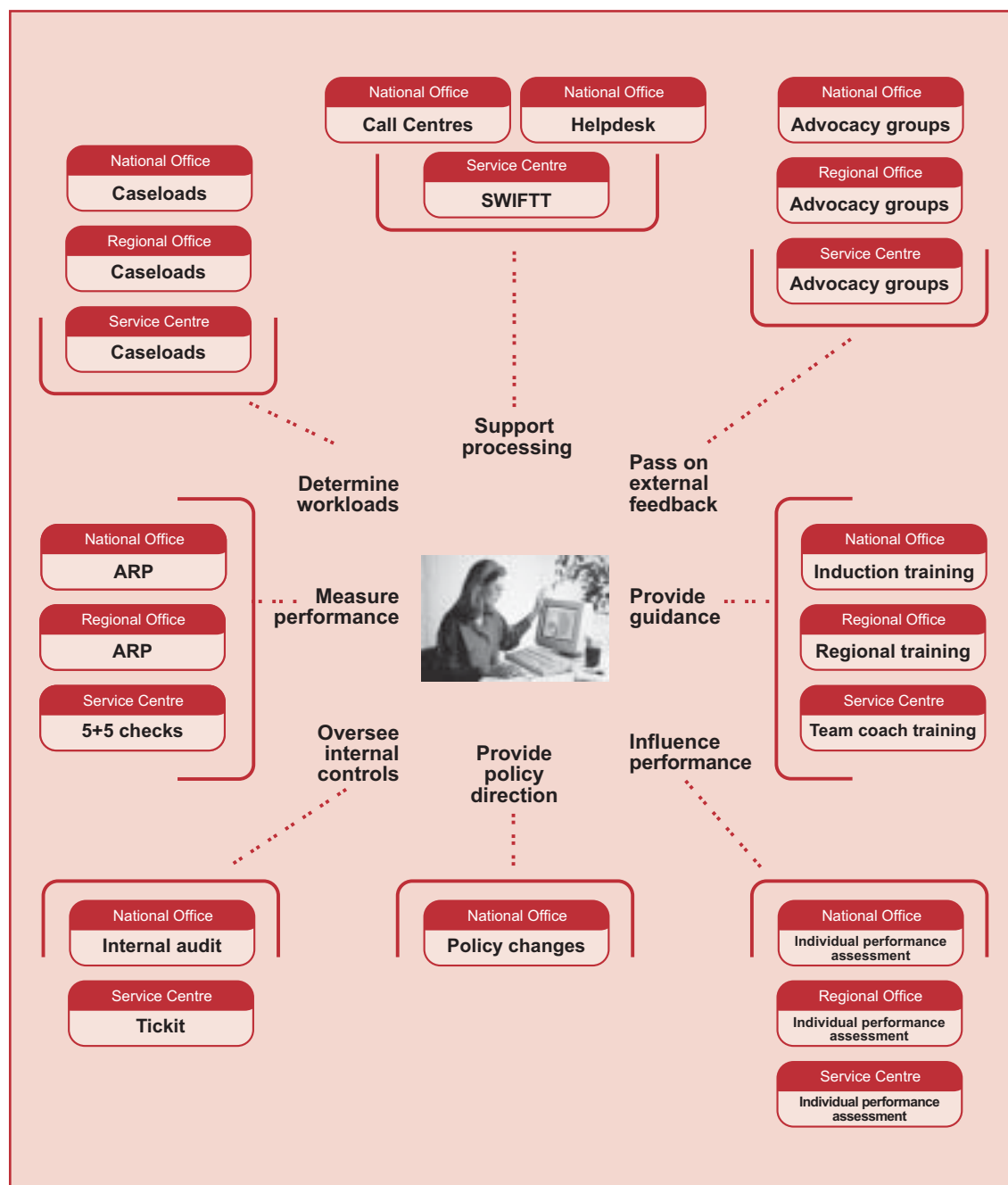
### *Which of These Activities Have the Greatest Effect on Accuracy?*

- 2.14 In our view, some of these activities, if altered, would have a noticeable effect on levels of accuracy. Others, while important, would have a less direct effect.
- 2.15 We concluded from our own observations and discussions with the Ministry's staff that the following four activities potentially have the greatest effect on case manager accuracy:
- case managers' caseloads;
  - checks on case managers' performance (the 5+5 checks);
  - team coaches' support and training of case managers; and
  - case manager performance assessments.
- 2.16 We explore some of the potential implications for accuracy of each of the four activities in Part Four on pages 41-53.



# ORGANISATION AND MANAGEMENT OF BENEFIT ADMINISTRATION

*Figure 6*  
*Activities That Can Affect Case Manager Accuracy*





## Part Three

# Systems and Processes for Ensuring Benefit Accuracy





### Introduction

- 3.1 In this part we describe the Ministry's benefit processing system and the processes for dealing with an application for benefit. We look at the circumstances that can lead to an inaccurate benefit payment.

### The Benefit Processing System (SWIFTT)

- 3.2 The computer system that processes the benefit information entered by case managers is known as SWIFTT (Social Welfare Information for Tomorrow Today). It was introduced in 1990. It is a screen-based system – it does not use stacked 'windows' that most computer users are familiar with today. SWIFTT is the biggest payments system in New Zealand, handling more than 2.5 million payment transactions a fortnight.
- 3.3 There is no evidence of SWIFTT directly causing benefits to be paid inaccurately. However, SWIFTT has some constraints that limit its use to identify inaccuracy, including:
- **Automatic checks to assist fraud detection** are an important way to reduce the number of inaccurate payments as a result of fraud, but have only recently been devised as part of entitlement checks. For example, through an application that has allowed the format of address data to be standardised, it is now possible to identify whether any other benefits are being paid into the bank account number supplied by a new applicant.
  - **Alerts to case managers** when they enter certain information in a way that has previously led to complications – SWIFTT will not alert them to check certain details or direct them to input additional information.
  - **Compulsory fields to explain decisions** – SWIFTT contains only voluntary fields in which case managers can note the reasons for making a decision. This can lead to "unverifiable" items (see paragraphs 5.18-5.23 on pages 61-63). A case manager's decision-making would be easier to follow if SWIFTT contained compulsory fields for reasons for decisions on selected critical aspects of an assessment.



## SYSTEMS AND PROCESSES FOR ENSURING BENEFIT ACCURACY



- **Management information** available from SWIFTT is limited. It provides information on over-payments and debt balances by individual client, but not the extent and causes of under- and over-payments. However, data from the SWIFTT, UCVII and TRACE systems is consolidated in the Ministry's Information Analysis Platform (IAP) and is used by the Ministry's Business Units to provide management information.

### *Conclusions*

- 3.4 SWIFTT, while old, is sound in terms of accurately processing benefits. It has some constraints that limit its use to identify accuracy. The Ministry has developed some new functions, and is currently formulating an Information Technology Strategy relating to the overall future development of the system.

#### **Recommendation 1**

We recommend that –

The Ministry continues to explore further enhancements to SWIFTT involving automatic checks, alerts to case managers, compulsory fields to provide explanations of decisions, and the production of management information.

## Applying for a Benefit and Supplying Information

### *Contact With the Ministry*

- 3.5 People often contact the Ministry to enquire about benefits before making formal application. Often, the enquiry is by telephone to a Ministry Contact Centre. Such contact does not constitute a claim or application for a benefit.



### Application and Assessment

- 3.6 What constitutes an “application” for a benefit is not defined in law or elsewhere. An applicant has to ... *furnish in such form as the [Social Security] Commission may require such personal and other particulars and (where applicable) such particulars of income and property as the Commission may require...*<sup>8</sup>
- 3.7 However, arising from the judgment in the *Scoble* case<sup>9</sup> a number of key expectations can be established:
- The Ministry should regard the Social Security Act as a “working document” so as to give it a “reasonable and practical application”. The Ministry should not, by being overly technical, deny an eligible person assistance, but at the same time it should operate in a way which is administratively efficient and does not waste public money.
  - The Ministry should make an intelligent appraisal of a person’s circumstances as presented to it; and, as necessary, enquire further to determine whether the person is eligible for a benefit appropriate to the circumstances and, if so, which benefit(s).
  - An application is a “claim for assistance to meet an identified need”, for a “particular purpose”. A request by a person for information is not enough to constitute an application.
  - The Ministry can treat an application for one benefit as an application for another benefit, so long as the applicant has sought assistance for a “particular purpose”; and the Ministry can do so either at the time of application or later.
  - The Ministry is required to inform a claimant of benefits relevant to the circumstances the claimant discloses or which the Ministry knows about.
- 3.8 When interviewing applicants, case managers seek further information. *Scoble* has reinforced the importance of this information-gathering process in order to assess eligibility for assistance. In addition, the Ministry periodically requires beneficiaries to provide updated information on their circumstances in the course of standard reviews of their benefits.

8 Regulation 7(1), Social Security (Monetary Benefit) Regulations 1971 (SR 1971/167).

9 *Chief Executive of the Department of Work and Income v Scoble* [2001] NZAR 1011.



## SYSTEMS AND PROCESSES FOR ENSURING BENEFIT ACCURACY

### *How Can a Benefit Be Inaccurate?*

- 3.9 In Figure 5 on page 29 we identified the accuracy risk areas in the benefit system. There are five main circumstances that can cause a benefit to be paid inaccurately:
- An applicant or beneficiary unwittingly provides incorrect information or unwittingly fails to provide complete information about their circumstances. For example, an applicant might omit to provide information about the number of people sharing the rental of a dwelling, thereby causing the wrong level of Accommodation Supplement to be paid.<sup>10</sup>
  - An applicant or beneficiary intentionally provides false information to the Ministry, thereby obtaining an amount of benefit to which they are not entitled. In such circumstances, the applicant may be guilty of fraud.<sup>11</sup>
  - A member of the Ministry's staff (usually a case manager) makes an incorrect decision about one or more of –
    - the person's eligibility;
    - their entitlement; and
    - the amount of benefit that is payable.
  - A member of the Ministry's staff makes a clerical error.
  - The failure of the Ministry's payment system causes an error.
- 3.10 Any of these circumstances – including fraud – can lead to the Ministry making an under-payment or an over-payment, or a payment that should not be made at all.

<sup>10</sup> The Ministry classifies these situations as "innocent over-payments".

<sup>11</sup> It is important that the term "fraud" is used accurately. There are some complex legal considerations to take into account in determining whether or not a benefit has been obtained "fraudulently" within the meaning of the Crimes Act 1961 and/or the Social Security Act 1964. In general terms, it is insufficient simply to establish that incorrect information was supplied to the Ministry, or even that the information was supplied in circumstances where the beneficiary knew that it was wrong.





## SYSTEMS AND PROCESSES FOR ENSURING BENEFIT ACCURACY

- 3.11 When receiving a benefit, there is a legal duty on the beneficiary to update the Ministry as soon as possible of any changes in their circumstances that are likely to affect their benefit entitlement. These can include changes in:
- with whom the beneficiary is living;
  - the beneficiary's level of earnings;
  - the number of dependants the beneficiary supports; or
  - what the beneficiary is paying for accommodation.
- 3.12 The Ministry has the discretion under law to review benefits at any time. In practice, standard review intervals are set – 26 weeks or 52 weeks – but it might be more cost-effective to match review intervals to the average period over which beneficiaries' circumstances undergo some material change. The Ministry does not collect information that would help it to determine such intervals.

### *People Who Do Not Supply Complete and Correct Information*

- 3.13 When a person does not supply complete or correct information, and subject to the expectations listed in paragraph 3.7, the person may not be entitled to any payment or may be entitled to payment of an amount different to that actually paid. Because the Ministry measures accuracy in terms of its own actions (consistent with the Purchase Agreement), it does not regard these incorrect payments as “inaccuracies” for the purposes of its Accuracy Reporting Programme, which we describe in Part Five.
- 3.14 As noted in paragraph 3.3, there are difficulties in extracting and analysing for management purposes SWIFTT data about the extent and causes of under- and over-payments. We consider that reliable information on cases where the information provided by an applicant was incomplete or incorrect could enable the Ministry to devise corrective actions. For example, if it were found that applicants often misunderstand a particular question or obligation, steps could be taken to clarify the instructions. This kind of information could prove to be particularly useful in helping to simplify the benefit process.



## SYSTEMS AND PROCESSES FOR ENSURING BENEFIT ACCURACY



### *Conclusions*

- 3.15** In summary, the Ministry's obligation is to pay benefits correctly on the basis of the information available to it (whether the information was provided by or obtained from the applicant/beneficiary, or was otherwise known to it). It measures accuracy in terms of this obligation.
- 3.16** Under- or over-payments that arise other than from a Ministry error (for example, because the applicant provides incorrect or incomplete information) can still have consequences for the Ministry or the beneficiary, or both, of the kind described in paragraphs 1.6-1.9. For example, where incomplete information leads to the beneficiary not getting a second-tier benefit. There are likely to be benefits in learning more about why under- or over-payments happen, with a view to devising ways to avoid them, or to identify them as soon as possible. In this respect, the Ministry has implemented a range of early intervention strategies.

#### **Recommendation 2**

We recommend that –

The Ministry continues to explore cost-effective ways of extracting and analysing data about the extent and causes of under- and over-payments, and uses the results of the analysis to devise ways to avoid under- and over-payments or to identify them as soon as possible.



## Part Four

# The Pivotal Role of the Case Manager





### Introduction

- 4.1 In paragraph 2.15 on page 30, we concluded that the following activities potentially have the greatest effect on case manager accuracy:
- case managers' caseloads;
  - checks on case managers' performance (the 5+5 checks);
  - team coaches' support and training of case managers; and
  - case manager performance assessments.
- 4.2 In our view, it is important that the Ministry understands the effect that these activities have on accuracy.
- 4.3 The Ministry has information about the volumes for each of these activities, and reports this information internally, although we found that some comparative volume information could be more systematically collected and circulated. However, the Ministry does not have information on the effects that these activities have upon accuracy.
- 4.4 This lack of information on effect extends beyond these activities – such as to policy development, which did not, for instance, consider the effect on accuracy when creating the new Domestic Purposes Benefit Programme in 2002. However, WINZ and the Ministry's Sector Policy Group are working closely together (under the Future Directions Programme) to simplify the benefit system. This will have an effect on accuracy.
- 4.5 In this part, we explore some of the potential implications for accuracy of each of the four activities.

### Case Managers' Caseloads

- 4.6 High caseloads can create pressures on case managers that may lead to lower levels of accuracy. This effect can be mitigated, such as by improving case managers' skills, or improving support systems to enable them to cope better. However, there is a limit to the volume of cases that an individual case manager can handle without putting accuracy at risk.



## THE PIVOTAL ROLE OF THE CASE MANAGER

4.7 Decisions on caseloads are made at three levels:

- national;
- regional; and
- service centre.

### *National Level Resource Allocation*

4.8 The Ministry distributes funding for Service Delivery in a standard manner across regions according to caseload. It uses a “pure” model for the distribution, in that most of the funding is allocated according to the number of clients<sup>12</sup> that each region has on its register. There are no substantial adjustments for other factors related to accuracy, and some of the adjustments that are made are also related to pure caseload (e.g. ratio of team coaches to case managers).

4.9 The Ministry believes that it is best to have almost all funding distributed to Regional Commissioners and for them to manage within their allocation. Therefore, only a very small amount – less than 0.5% – is retained by National Office as a ‘bidding fund’ that would potentially be available to deal with short-term pressures in particular regions.

4.10 We concluded that regional variations could place pressure on accuracy in a number of ways that affected caseloads. For example, a region could have a disproportionate number of beneficiaries with complex needs requiring assessments for different benefit types that can be particularly time-consuming for case managers.

4.11 Using the number of applications that a region receives for Special Needs Grants and Advances as an indicator of the complexity<sup>13</sup> of beneficiaries’ needs, we found that there are large differences between regions (see Figure 7 on the opposite page). The lowest proportion of applications for a Special Needs Grant was only 52% of the number of working-age beneficiaries, and the highest proportion of applications was 136%.

<sup>12</sup> This is greater than the number of beneficiaries, because it includes people not on a benefit whom the Ministry is helping to find work.

<sup>13</sup> We acknowledge that this is a “rough” measure. To be more precise, it would be necessary to separate out regional differences in eligibility from regional differences in the interpretation of eligibility.





## THE PIVOTAL ROLE OF THE CASE MANAGER

*Figure 7*  
*Applications for Additional Assistance 1999 to 2002*  
*as a Proportion of Working-age Beneficiaries*

|                      | <b><i>Lowest Regional<br/>Ratio %</i></b> | <b><i>Highest Regional<br/>Ratio %</i></b> |
|----------------------|---|--|
| Special Needs Grants | 52  | 136  |
| Advances             | 41  | 91   |

- 4.12 We concluded that the average complexity of benefits influences, but does not uniquely determine, a region's accuracy rate. Because there are a number of factors that influence overall accuracy, it is possible that a region can have a complex population of beneficiaries and still record high accuracy figures by intensifying other activities. However, because the Ministry does not know the effect on accuracy of individual activities, it would be difficult to determine whether and how a region's accuracy figure has been achieved.

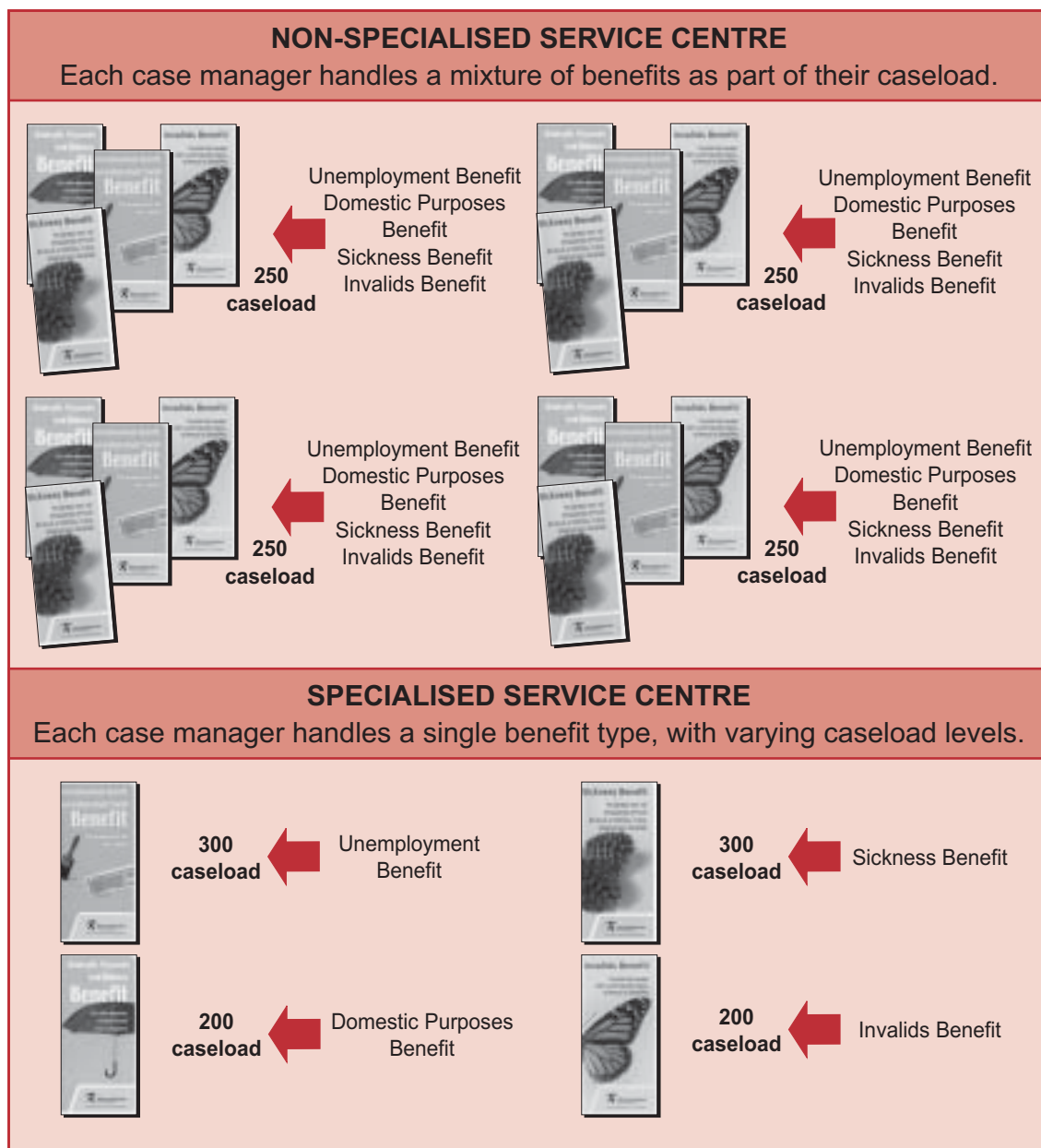
### *Caseload Decisions at Regional Level*

- 4.13 Within the national funding allocated to them, Regional Commissioners are free to set caseloads for particular service centres as they see fit – they are not expected to stick rigidly with the national caseloads on which the funding was based. The main constraint on Regional Commissioners is the expectation that they will meet KPI targets – so their decisions on where to place resources need to be made with this in mind.
- 4.14 For example, Regional Commissioners may vary:
- the degree of specialisation in their service centres – Figure 8 on the next page illustrates the two extremes of specialisation and non-specialisation; and
  - the proportion of investment in team coaches (who support and train case managers – see paragraphs 4.36-4.44) – which may need to be greater than the proportion indicated in the national model – for example, in an area that has relatively high staff turnover and high proportions of relatively new staff.



## THE PIVOTAL ROLE OF THE CASE MANAGER

Figure 8  
Options for Managing Caseloads – To Specialise or Not?



- 4.15 In the Wellington Region we found that some of the savings from having a high turnover and relatively high numbers of new staff starting had been used to establish a new post at regional level to concentrate on risk management – partly in order to address having high numbers of new and relatively inexperienced staff.





## THE PIVOTAL ROLE OF THE CASE MANAGER

### *Caseload Decisions at Service Centre Level*

- 4.16 Each service centre manager can also flexibly deploy staff. Variations from the regional pattern must be approved by the Regional Operations Manager to ensure that any changes are consistent with the region's goals.
- 4.17 Variations at service centre level are usually adjustments to caseloads to take account of case manager experience – for example, new case managers are often given lighter caseloads until they attain the necessary level of skill. Types of work allocated may also be varied – for example, a member of staff who is considered to be efficient, accurate, and conscientious in decision-making and recording might deal with a disproportionate number of the “once-only” changes to circumstances that beneficiaries report to the service centre.

### *Conclusions*

- 4.18 The Ministry's arrangements for allocating resources and devolving responsibility for managing them are consistent with giving managers appropriate responsibility for what they control and making them accountable for results.
- 4.19 There is limited sharing of information between regions. Benefit administration is likely to be improved if reliable assessments of the benefits and risks to accuracy of different approaches to improving accuracy were to be routinely shared with other regions rather than remaining as lessons held within particular regions.
- 4.20 Because accuracy results are only reported annually to the national level, any region that adopts an approach that turns out to alter accuracy will not come to national management's attention in a timely way. In the meantime, if the impact is to reduce accuracy, beneficiaries may be disadvantaged for a considerable length of time.
- 4.21 Regions are resourced according to the number of clients. Those regions with a greater than average number of beneficiaries with complex circumstances could be disadvantaged, because the additional work arising from dealing with these beneficiaries could leave them less able than other regions to meet their accuracy target.



## THE PIVOTAL ROLE OF THE CASE MANAGER

### Recommendation 3

We recommend that –

The Ministry continues to promote sharing of information among regions and with National Office on approaches to staff deployment and managing caseloads.

### Recommendation 4

We recommend that –

For those regions that have a high number of beneficiaries with complex circumstances, the Ministry explores whether a relatively higher allocation to reflect the workload associated with complexity is justified.

## “5+5 Checks” on Case Managers’ Performance

- 4.22 Case managers’ pay and promotion are linked to the achievement of individual targets that are in turn linked to those set at regional level.
- 4.23 One of the ways in which the Ministry monitors case managers’ individual performance on accuracy is through “5+5 checks”. These involve the team coach checking the accuracy of five applications and five reviews completed by the case manager within the last month. The team coach has discretion to undertake additional checks to ensure that the five applications and five reviews originally chosen are not unrepresentative of the case manager’s work.
- 4.24 For a new case manager, the team coach checks every application and review until the case manager reaches a prescribed level of proficiency, after which the 5+5 checks are applied. If the case manager’s accuracy deteriorates at any point, the number of checks is increased until the service centre manager is satisfied that the problem has been rectified and the case manager has returned to the prescribed level of proficiency.
- 4.25 As a minimum, the 5+5 checks follow those required for the Accuracy Reporting Programme (“ARP” – see Part Five), and in some regions also involve wider checks to cover the general work practices of case managers.





## THE PIVOTAL ROLE OF THE CASE MANAGER

- 4.26 Our technical referee for this audit identified some significant inference risks when 5+5 checks are used to assess the performance of individual case managers. The risks derive from the very small sample taken of each case manager's work. The problem is manageable but, in our view, does need to be addressed.

### *How Do 5+5 Checks Assist Accuracy?*

- 4.27 5+5 checks assist accuracy in two main ways:
- by providing the information required to measure individual case manager accuracy and to identify any need for improvements; and
  - by providing the same information for the same purposes at service centre level to enable managers to monitor and improve the accuracy of processing within the centre or within groups of case managers.
- 4.28 Because 5+5 checks are continuous, they enable the service centre manager to monitor performance throughout the year. By identifying trends in 5+5 check results, the manager can identify training needs or raise awareness of processing aspects among case managers and their team coaches.
- 4.29 At a large service centre, the number of 5+5 checks might be about 150 a month. The same centre would probably be asked to check only four or five applications or reviews for the Accuracy Reporting Programme.
- 4.30 The Ministry does not use 5+5 checks for national reporting, because they are less objective and less consistent than the checks made under the ARP. Not all regions review the results between service centres, and there is no review of the results between regions or between similar service centres in different regions.
- 4.31 After the bulk of our field work was carried out, the Ministry informed us that it has initiated a procedure of compiling Team Coach data (5+5 checks) from time to time to provide context or elaboration to nationally compiled data and to inform other quality improvement mechanisms such as Helpline or Quality Control.



## THE PIVOTAL ROLE OF THE CASE MANAGER

- 4.32 If ARP and 5+5 data could be validly “combined”, a better picture of performance at regional level could be obtained. One method by which the ARP and 5+5 sampling systems might be linked involves what is called ‘double sampling’. In essence, the method involves using a special sample to determine the nature of the relationship between measurements of accuracy derived from ARP and 5+5 approaches and using this relationship to infer regional accuracy information from 5+5 data.

### *Conclusions*

- 4.33 KPIs are a central feature of the Ministry’s culture. Case managers have clear, measurable targets to aim for in their day-to-day work.
- 4.34 The 5+5 checks provide valuable information on accuracy in service centres, and are an effective means of achieving and maintaining accuracy. In those regions that look at the results across service centres, they also provide useful comparative information.
- 4.35 The lack of more high-level comparative analysis of the results – particularly between service centres dealing with similar beneficiary profiles and problems – is a missed opportunity to compare performance and to transfer lessons.

#### **Recommendation 5**

We recommend that –

The Ministry investigates the possibility of using aggregated 5+5 data at regional level to draw inferences about regional performance through the use of appropriate statistical techniques such as double sampling.





## THE PIVOTAL ROLE OF THE CASE MANAGER

### Role of Team Coaches

- 4.36 The role of team coaches is wider and more supportive than just checking case managers' work. For example, a case manager can book a time with the team coach to discuss a specific problem. Team coaches also identify shortcomings through the 5+5 checks, and work one-on-one with the case manager to help improve their performance.
- 4.37 Team coaches can also manage training of groups at service centre level (held on most Wednesdays throughout the Ministry). This includes training to bring in new Ministry-wide developments as well as issues specific to the service centre – which means that the team coaches generally have a good understanding of Ministry-wide changes.
- 4.38 All the team coaches we spoke to saw their role as fostering a culture of continuous improvement in their service centres. Team coaches are generally experienced case managers, so they understand the pressures and difficulties that case managers face and have the authority and experience to gain respect.
- 4.39 Staff throughout the Ministry – case managers, service centre managers, and Regional Commissioners – acknowledge the importance of team coaches in achieving and maintaining accuracy. Case managers we spoke to praised the concept of team coaches, and said they valued the support and guidance they provide.

### *How Many Team Coaches Are Required?*

- 4.40 The Ministry's national funding model provides for a ratio of one team coach to every 25 case managers – a ratio that was set in 1995 after a review of the numbers needed to maintain targeted performance. In 2000 the Ministry re-allocated funds to ensure that the ratio did not drop below this level.
- 4.41 Regional Commissioners can alter this ratio as they see fit. Some regions have reduced the number of team coaches and increased the number of case managers – enabling them to marginally reduce the caseload per case manager. Other regions have done the reverse – increased the number of team coaches, funded by a reduction in case managers and higher caseloads.



## THE PIVOTAL ROLE OF THE CASE MANAGER

- 4.42 In addition, the role that a team coach performs can be varied – such as to only undertake checking or training, or to take on other administrative tasks.
- 4.43 We found that, before 2002, the Ministry did not seek to collect information on different practices in relation to numbers and the role of team coaches. Experience during 2002 in some regions suggested that varying the number of team coaches could be counterproductive and this information was shared among all regions.

### *Conclusion*

- 4.44 Team coaches provide important and responsive support and training to the case managers. We consider that not sharing information on how regions use team coaches is (as with the 5+5 checks) a missed opportunity to compare performance and to transfer lessons.

#### **Recommendation 6**

We recommend that –

The Ministry continues to promote sharing of information among regions and with National Office on approaches to the use of team coaches.

### **Case Manager Performance Assessment**

- 4.45 The KPI for processing accuracy appears throughout the Ministry – right down to individual case managers' performance agreements. Each case manager is also expected to meet a large number of other performance standards as part of their performance agreement. The majority of these other performance standards are directed at the employment element of the case manager's duties – for instance, the number of long-term unemployed people the case manager has placed in work.





## THE PIVOTAL ROLE OF THE CASE MANAGER

- 4.46 We found variation between regions in how performance pay is related to accuracy. Some regions see an accuracy standard as a compulsory target – a case manager must achieve or better it before any bonus will be paid. The aim is to ensure that, no matter which case manager handles a beneficiary's case, a minimum level of accuracy can be expected.
- 4.47 Other regions assess performance more broadly across all the performance standards. A case manager might have a low accuracy result (below the 90% level at which the Ministry defines a case manager as competent) but still receive a bonus if they have performed at a high level in respect of employment targets.

### *Conclusion*

- 4.48 The Ministry has a systematic approach to linking individual performance of its case managers to its KPI for processing accuracy. However, the way some regions assess performance creates a risk that staff will not see accuracy as a high priority, potentially leading to an inconsistent level of accuracy between case managers and regions. We do note that the Ministry audits performance assessments from time to time and we encourage this to continue.

#### **Recommendation 7**

We recommend that –

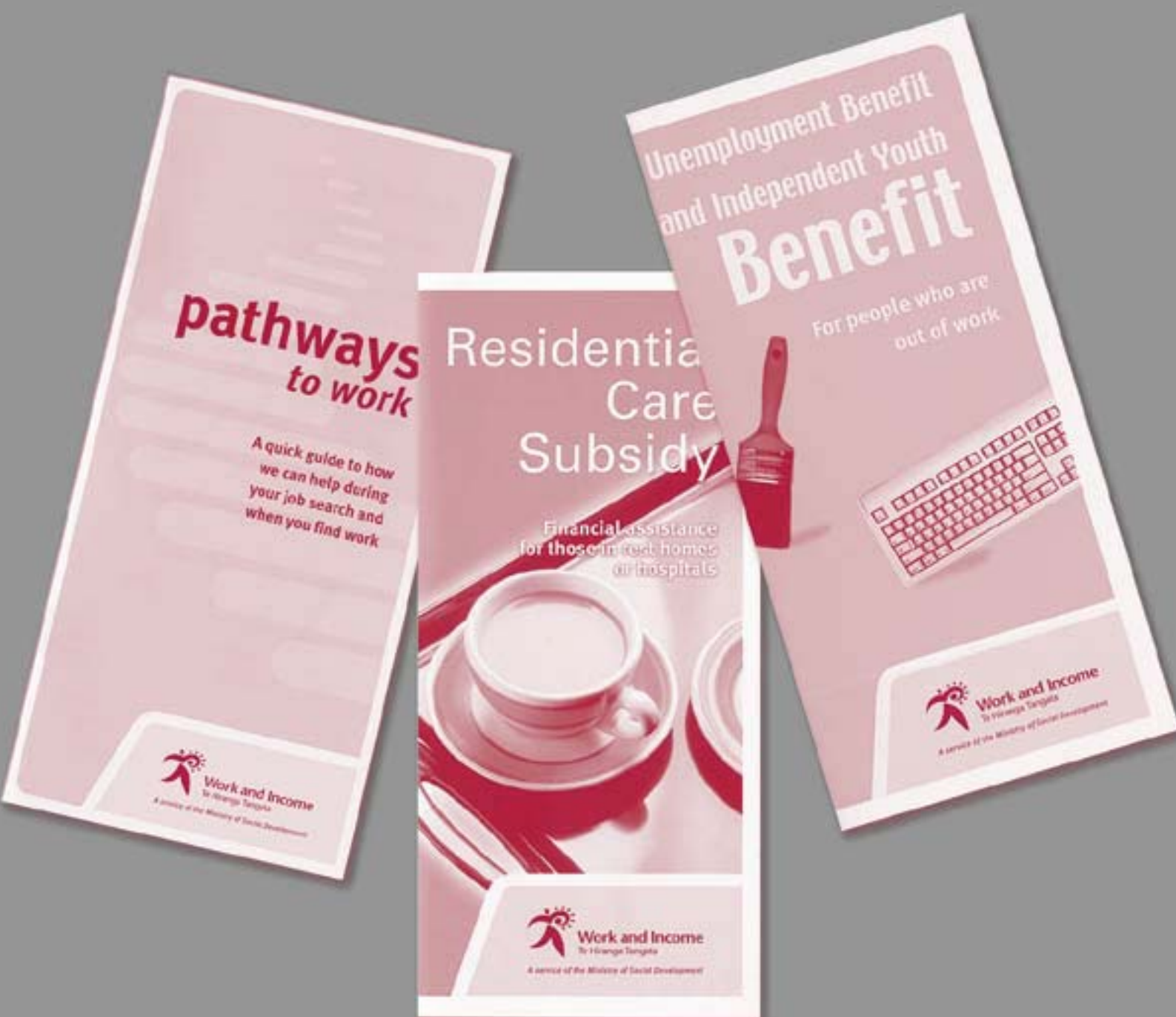
The Ministry requires all regions to assess the performance of staff consistently. The chosen method should provide an incentive to staff to accord accuracy an appropriately high priority.





## Part Five

# Accuracy Measurement and Reporting







## Defining Accuracy

- 5.1 An observer who had the luxury of perfect information might define perfect accuracy in this context as *the condition where everyone in New Zealand who has a legal entitlement to a benefit receives that exact entitlement on the date when it is legally due*. However, such perfect accuracy cannot be achieved.
- 5.2 The Ministry has designed its accuracy measure to reflect the legal position that it cannot grant a benefit until application is made for it (see paragraph 2.9 on page 29).
- 5.3 In this part, we look at how the Ministry measures and reports accuracy through its Accuracy Reporting Programme.

## The Ministry's Performance Measure for Accuracy

- 5.4 The performance measure in the Ministry's Statement of Intent for accuracy of benefits granted to working age beneficiaries is that *the percentage of entitlement assessments completed accurately will be no less than 88%-90% (92% for people receiving New Zealand Superannuation)*. This is the only performance measure of accuracy that is reported to Parliament.
- 5.5 The Ministry's prime focus is thus on accuracy in relation to eligible people who have applied for a benefit. The Ministry's Accuracy Reporting Programme (ARP) does not address eligible people who have either:
  - made no contact with the Ministry; or
  - contacted the Ministry but not applied for a benefit.
- 5.6 The Ministry's approach to accuracy is consistent with its legal obligation (see paragraphs 2.9-2.10 on page 29). The Purchase Agreement – in which the Ministry's performance on accuracy is limited to matters after the application has been received – specifically refers to the accuracy of the *processing of applications and reviews<sup>14</sup> of benefit entitlement*.

14 A "review" is essentially a decision to change the benefit as the result of a change in any aspect of the beneficiary's circumstances – for example, to change the amount of accommodation supplement payable because of a change in the beneficiary's accommodation costs.



## ACCURACY MEASUREMENT AND REPORTING

### The Accuracy Reporting Programme – What It Measures and What It Doesn't

- 5.7 The ARP measures and reports the accuracy performance measure against an annual target stated in the Output Agreement. The Ministry samples about 250 applications and 250 reviews a year from each of its 13 regions, and checks that these have been processed correctly. Currently, the target is that no less than 88-90% of cases sampled for working age beneficiaries will be completed accurately (92% for people receiving New Zealand Superannuation).
- 5.8 The target has varied over time. In 1996-97, the Department of Social Welfare's performance target for its Income Support Service was 95% accuracy for applications and 95% for reviews. In that year, the Income Support Service did not meet the target – actual performance was 82% for applications and only 74% for reviews.<sup>15</sup> An accuracy target of 80% was set in 1998-99 and has since increased incrementally to the present level as performance has improved.
- 5.9 The Ministry's reported accuracy performance for 2002-03 was 89.2% (2001-02 – 91.1%) for working-age beneficiaries, and 93.1% (2001-02 – 96.6%) for people receiving New Zealand Superannuation. The Ministry considers that the accuracy target (88-90% for working-age beneficiaries and 92% for superannuitants) is at a level that successive governments have been comfortable with. It has therefore not recently sought to make any fundamental changes to accuracy targets.

### *Scope of the Accuracy Reporting Programme*

- 5.10 Since the scope of the ARP covers only the Ministry's performance in assessing benefit applications, it does not and cannot estimate the proportion of eligible beneficiaries who have not "applied". In the Ministry's terminology, this is a matter of "uptake" rather than "accuracy".
- 5.11 At an operational level, an application results in the creation of a computer record relating to that person and that transaction. It is from these computer records that the Ministry draws its sample to measure and report on the accuracy of benefit payments.

<sup>15</sup> At that time, the accuracy rates of new applications and reviews of existing benefits were reported separately.





- 5.12 Consequently if, for example, someone –
- does not know about their possible eligibility (and so does not contact the Ministry); or
  - knows about their possible eligibility but makes a deliberate choice not to contact the Ministry; or
  - knows about their possible eligibility and contacts the Ministry but receives incorrect advice that causes them not to make an application –
- they will not be included in the sample.

### *People Who Are Eligible But Not Aware of It*

- 5.13 An example of a person who is eligible for a benefit but not aware of it is someone on a low income (and not receiving any first-tier benefit) who may be eligible for an Accommodation Supplement (a second-tier benefit) to offset their housing costs. The person may mistakenly believe that they could receive the Accommodation Supplement only if they are receiving one of the first-tier benefits. Their lack of awareness of eligibility results in them bearing the full cost of their housing.
- 5.14 Because the Ministry considers these cases as people who do not take up the benefit rather than people who are being underpaid, it does not regularly collect information on:
- the likely size of the group;
  - the characteristics of the group; or
  - the average dollar amount that each person in the group should be receiving.

In making this point, we also acknowledge the Ministry's substantial and ground-breaking work in relation to the Living Standards Research Project. This is an ongoing research programme focusing on developing a comprehensive description of the living standards of New Zealanders, which will enable governments and communities to develop evidence-based policies to address disparities between different groups of New Zealanders.



## ACCURACY MEASUREMENT AND REPORTING

**5.15** It is feasible to run a special exercise to collect information that could not be obtained through the ARP. Figure 9 below shows the results of a national survey the Ministry conducted in 1996 to ascertain the proportions of people who were and were not receiving the Accommodation Supplement.<sup>16</sup> The three key features of the results are that:

- Of the people determined to be eligible for the supplement (19% of the sample), just under two-thirds were receiving it.
- Of the people determined not to be eligible (77% of the sample), a small group was receiving the benefit. The people in that group had (presumably) been granted the benefit on the basis of incorrect information (innocently or falsely given in their application) or on the basis of correct information incorrectly assessed by the Ministry.
- The eligibility of 4% of the people in the sample could not be determined. Inadequate or incomplete information creates the risk that a person eligible for a benefit might be denied it or a person not eligible for a benefit might be granted it.

**Figure 9**  
*Benefit Accuracy Information – 1996 National Survey of Eligibility for the Accommodation Supplement*

|   | Proportion of sample % |
|---|------------------------|
| People eligible for the supplement –            |                        |
| • receiving it                                  | 12                     |
| • <b>not</b> receiving it                       | 7                      |
|   | <hr/>                  |
|   | <b>19</b>              |
| People <b>not</b> eligible for the supplement – |                        |
| • receiving it                                  | 3                      |
| • <b>not</b> receiving it                       | 74                     |
|   | <hr/>                  |
|   | <b>77</b>              |
| Indeterminate eligibility                       | <b>4</b>               |
|   | <hr/>                  |
|   | <b>100</b>             |

<sup>16</sup> The survey was of people in both rental accommodation and their own homes after the 1993 Government decision that *accommodation assistance (through a second-tier benefit called the Accommodation Supplement) should be available to low-income families, whether beneficiaries or wage earners, to assist with either rental or home-ownership costs.*



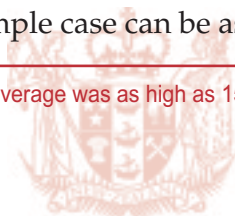
## Conclusions

- 5.16 The ARP measures and reports accuracy on the basis of the Ministry's statutory obligations and Purchase Agreement. It does not, and is not intended to, provide a wider measure of benefits being paid accurately to all who are eligible to apply.
- 5.17 However, it is possible to measure and report data not collected by the ARP by conducting separate special exercises in data collection and analysis.

## Unverifiable Items

- 5.18 Some of the cases selected as part of the ARP sample cannot be assessed as correct or incorrect because files or parts of files are found to be missing. The Ministry excludes these cases – which it refers to as “unverifiabiles” – from the ARP sample results. At the time the decision to exclude unverifiabiles from the sample results was taken, we agreed to this approach. However, we now take the view that the decision was incorrect, because cases for which papers cannot be found may well have a different error rate from the rate for cases that are well documented. Excluding them may therefore result in incorrect estimates.
- 5.19 Figure 10 on the next page the regional accuracy results against the national target for working age beneficiaries, excluding unverifiabiles, as measured by ARP.
- 5.20 The national average for unverifiabiles as a proportion of the sample was 6.6% in 2001-02.<sup>17</sup> The Ministry has been trying to reduce the level of unverifiable cases. The level of unverifiabiles has come down to 4% for the 2002-03 year.
- 5.21 We understand that the improved results have been achieved mainly by the Ministry putting in place specific strategies to ensure that the level of unverifiabiles was reduced in 2002-03. Since July 2002, the Ministry has:
- Introduced a procedure of sending, on the fourth working day of each month, a summary spreadsheet to each regional ARP liaison person providing the results of the sample and highlighting any outstanding unverifiable cases. The regional liaison person then signs off on the unverifiable spreadsheet and comments on each case referred back to them. They confirm how each case has been addressed with the case manager and advise ARP when additional information will be provided so that the original sample case can be assessed.

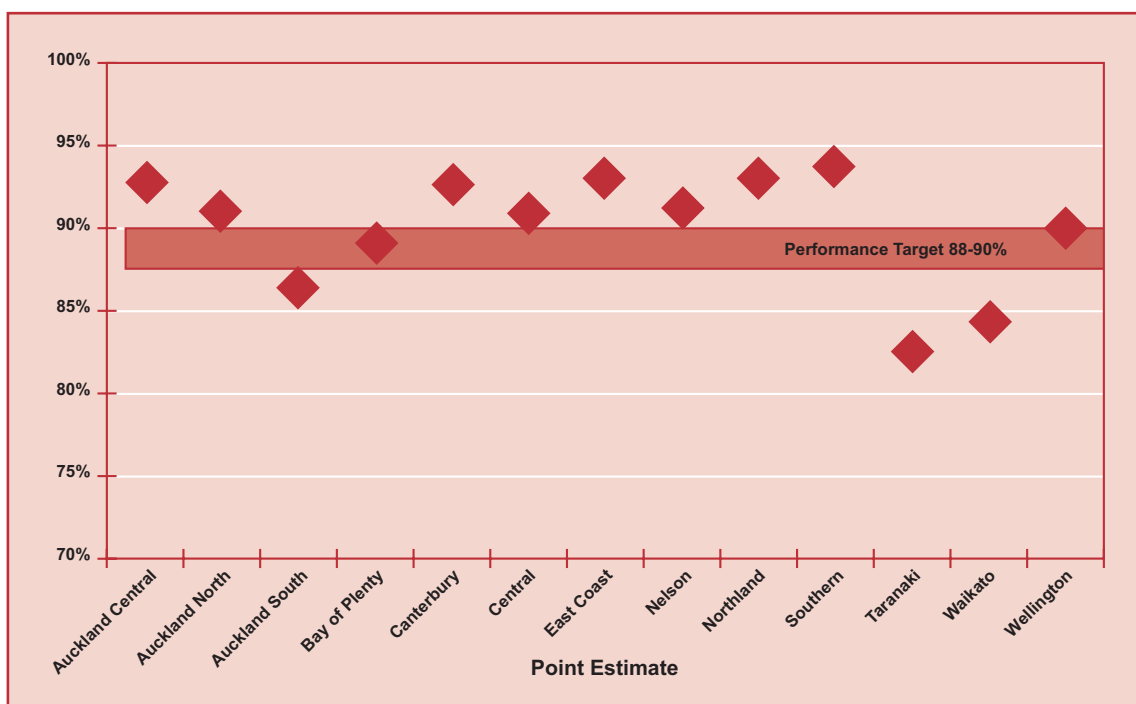
<sup>17</sup> In previous years the national average was as high as 15%.



## ACCURACY MEASUREMENT AND REPORTING

- Introduced a revised deadline from 12 months to 30 days to have cases returned for assessment from 1 July 2002.
- In those cases where no papers whatever could be located, introduced a process by which Benefit Control would check for indicators of fraud.
- Reviewed results to identify common types of errors to identify training issues and provided this information to regional liaison people and National Training and Helpline teams.

*Figure 10  
Regional Accuracy Results for 2001-02 Against the National Target as Measured by ARP Excluding Unverifiabiles*





## Conclusions

- 5.22 With the present practice of excluding unverifiabiles from the sample, the results could over-estimate the overall level of accuracy. By changing this practice to include unverifiable cases as errors, there would be a strong incentive for the Ministry to reduce the unverifiabiles to as low a level as possible. However, there would also be a risk that the results could under-estimate the overall level of accuracy, since it is unlikely that all unverifiable items contain errors.
- 5.23 An alternative approach would be for the Ministry to clearly explain unverifiabiles and disclose their level in its annual report.

### Recommendation 8

We recommend that –

The Ministry explains unverifiabiles and discloses their level in its annual report. The precise form of the disclosure should be agreed with the Audit Office as part of the annual audit.

## Uncertainty of ARP Results

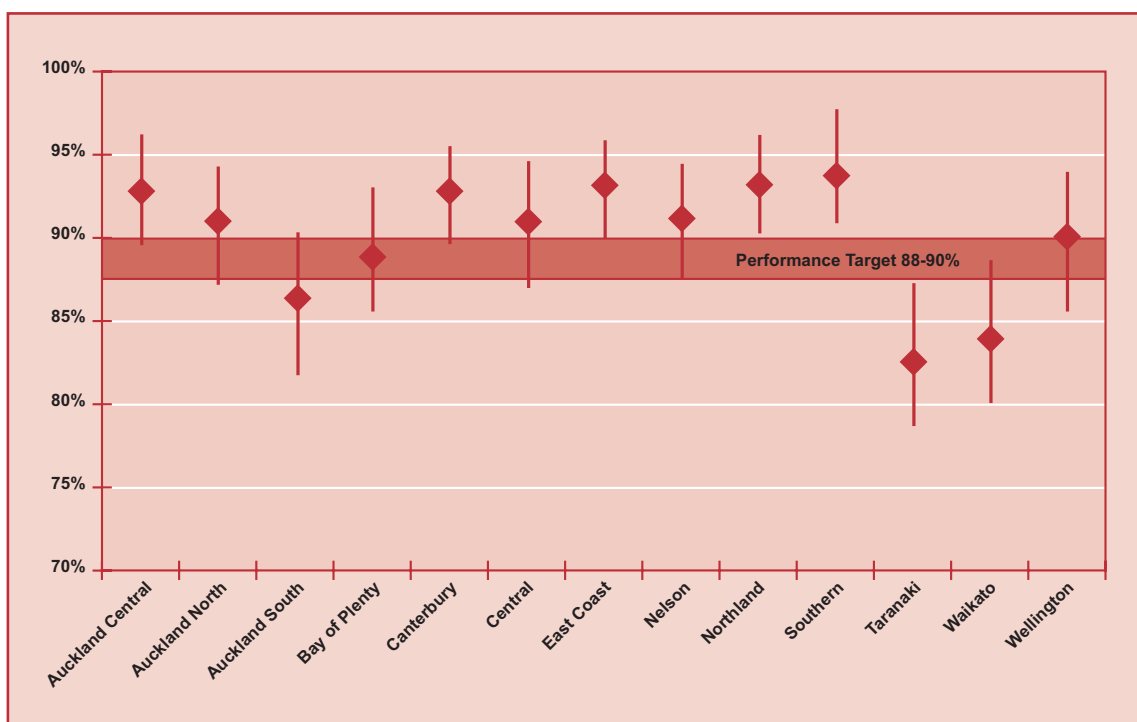
- 5.24 Subject to appropriate treatment of unverifiabiles, the ARP provides a suitable national measure – its sample size was designed to produce, at national level, a 95% confidence interval of  $\pm 1.4\%$  for both applications and reviews of existing benefits. When applications and reviews are aggregated (as at present), the 95% confidence interval for the national figure is  $\pm 1\%$ . This means that, if the point estimate (i.e. the result) from the sample was 90% accuracy, the range of accuracy would be between 89% and 91% (i.e.  $90\% \pm 1\%$ ).
- 5.25 When applied at regional level, the current ARP sample results in a much wider 95% confidence interval of about  $\pm 5\%$ . So, for example, while the Waikato region reported an accuracy level of 84.4% for 2001-02, the range of accuracy would be between 80.1% and 88.7%.



## ACCURACY MEASUREMENT AND REPORTING

- 5.26** This high degree of sampling error makes it difficult to draw useful conclusions from the analysis of ARP data at a regional level. For example, Figure 11 (below) illustrates that at a 95% confidence level there is only one region whose range of possible results does not either exceed the national target or overlap it. This means that it would be unsafe to conclude (except for the one region's case) that any particular region does not meet the national target.
- 5.27** Regional trend data is also difficult to interpret. For example, if a region produced a point estimate of 90% one year and 85% the next, it would be unsafe to conclude that there had been any change in performance between the two years, notwithstanding the difference between the two estimates. This is because the range that could be inferred from the first estimate would be 85% to 95% (i.e.  $90\% \pm 5\%$ ).
- 5.28** National Office provides the regions with their accuracy results, but only as a point estimate, as shown for each region in Figure 10 on page 62. Regions are not told the ranges of results, as shown in Figure 11 below.

*Figure 11  
Ranges of Regional Accuracy Results for 2001-02  
Against the National Target*





## ACCURACY MEASUREMENT AND REPORTING

- 5.29 Regions use their ARP results as one among a number of indicators of their performance in processing accuracy, and they take them into account in making operational changes to service delivery. Some regions aggregate their service centres' 5+5 checks as an additional regional measure of accuracy. We believe there is merit in this approach, so long as the regions are carrying out checks to ensure that the data from different service centres and teams is consistent.
- 5.30 The Ministry outlined to us a range of other information sources available to Regional Commissioners and their staff that can supplement ARP data in helping them form a view of their region's performance in respect of benefit accuracy. These sources include:
- The Ministry's risk management framework (Tickit), which requires service centre managers to certify compliance with key procedures.
  - Work product and process monitoring by a national quality assurance team to ensure consistency of practice across all regions.
  - Periodic reports by Internal Audit that are completed after examinations of both samples of work and operating procedures.
  - A system which provides a central repository for information training material and processes to help staff in dealing with frequently asked questions and requests, enabling them to provide consistent information regardless of where they are based. The system incorporates a number of tools, including one that lodges details of client requests for a review and monitors each review from the date on which the request was received.
  - Reports from the Ministry's Helpline service that record both the number of calls received from staff and the reasons for the call.
  - Reports on staff turnover rates and average duration of employment.
  - Individual performance assessments for each member of staff.
  - Information provided by regional training co-ordinators and the team coaches located in each service centre.
- 5.31 These additional sources do not provide statistically reliable measures of a region's performance, but they do give Regional Commissioners:
- information relevant to the competence and capability of the service centre staff in their regions;



## ACCURACY MEASUREMENT AND REPORTING

- assurance about the degree to which their staff are complying with established procedures; and
- supplementary indicators that would help to disclose any significant problems relating to processing accuracy – and provide a basis for reconsidering any potentially misleading picture produced by ARP results.

- 5.32 To the extent that ARP does not provide a precise measure of accuracy at regional level, that deficiency is partially compensated by the existence of these other indicators of performance. The Ministry expressed confidence that Regional Commissioners would not react to only one indicator but would form their views about the performance of their regions having regard to all the available information.
- 5.33 Because the Ministry has adopted a regional management structure and has delegated to Regional Commissioners the task of deploying their resources to manage performance effectively within their regions, we regard the quality of the management information available to Regional Commissioners as very important.
- 5.34 The Ministry considers that the Government does not require any region to stay within a specified range of accuracy so long as the national target is met. However, even if the Ministry is meeting performance targets at a national level, substandard performance by one or more regions could still adversely affect a considerable number of people. For example, beneficiaries in one region being disadvantaged where a low level of accuracy in their region leads to an above average level of under- and over-payments, and the associated consequences.
- 5.35 We agree that the additional indicators reduce the risk of relying on ARP alone. However, in our view, it is questionable whether the information currently available to regional managers is sufficient.
- 5.36 For example, if a Regional Commissioner were to be confronted with an apparently adverse change in the ARP results, in circumstances where staff turnover was higher than usual and there was some evidence of non-compliance with established procedures, any conclusion about the actual state of regional performance would still be uncertain. This is because the apparently adverse change in ARP might be wholly caused by sampling error.





## ACCURACY MEASUREMENT AND REPORTING

- 5.37 Conversely, an apparently positive ARP result that indicated no change or a positive change in the region's performance might be similarly misleading. When information is imprecise, there is always a risk of responding to false signals.
- 5.38 In the same vein, the Ministry's National Office needs to consider carefully the implications of the information it receives about regional performance and what management signals it sends in relation to its interpretation of that information.
- 5.39 National Office produces each month a document entitled Regional Performance Summary that contains a range of KPI indicators of regional performance. Each KPI is assigned one of four colours – blue, green, yellow or red – depending on whether the KPI exceeds, meets, is just below, or is significantly below a national performance standard. In response, Regional Commissioners each produce monthly for the National Commissioner a one-page National Overview Report that, among other items, provides commentary on the region's KPI results and indicates what corrective action is being taken.
- 5.40 In respect of ARP data, the Regional Performance Summary presents year-to-date point estimates for each region. Regions for which the point estimate is below the national target are signified in yellow or red, irrespective of whether the confidence interval overlaps the national target. In our view, this is likely to induce errors of inference.

### *Conclusions*

- 5.41 The ARP was not designed as a statistically valid instrument for measuring regional performance. Consequently, because the ARP estimates the level of accuracy of the total "population" of benefits on the basis of the accuracy of a relatively small random sample of those benefits, any use of ARP results in respect of regions needs to be approached with caution. Once down to regional level, the samples are small, and the smaller the sample on which an estimate is based, the greater the level of uncertainty of the estimate.
- 5.42 Earlier in the report (paragraph 4.32 on page 50) we pointed to the possibility of using 5+5 data to provide a better picture of accuracy performance at regional level. The foregoing discussion in this section highlights the importance of the issue.



## ACCURACY MEASUREMENT AND REPORTING

- 5.43 In our view, the relative levels of accuracy between different parts of the country are important. Regions' ARP results provide an imperfect measure of accuracy, but, in the absence of something better, it is acceptable for regional managers to use them as one of their operational tools. However, it is most important that, in doing so, they are clear about the uncertainties of the data they are using.

### Recommendation 9

We recommend that –

The Ministry continues to give the regions their ARP results, but in a form similar to Figure 11 on page 64 – showing each region's data at a 95% confidence level, and comparative data of other regions.

### Recommendation 10

We recommend that –

The Ministry provides all Regional Commissioners and Regional Operations Managers with training on the nature of sampling error and the appropriate interpretation of statistical estimates that include confidence intervals.

## Measuring and Reporting the Size of Errors and Fraud

- 5.44 The ARP provides a measure of the proportion of cases found to be accurate rather than any measure of the size of any incorrect payments or fraud. It cannot therefore be used to estimate the total amounts of under- or over-payments or fraud that potentially exist in the benefit system.
- 5.45 Benefit Control reports only the over-payments (including frauds) that it identifies and seeks to recover. It does not estimate the total value of over-payments (including fraudulent over-payments) that may actually exist.
- 5.46 It is nevertheless possible to assess the extent of under- and over-payments or fraud. In December 2001, the Ministry's internal auditors completed a study of the processing of benefit applications. The study found that 2.7% of the cases sampled had errors that resulted in benefits being paid inaccurately. However, it did not examine whether these payments were ongoing (i.e. would continue to be inaccurate week after week) or were once-only errors.





## ACCURACY MEASUREMENT AND REPORTING

- 5.47 In practice, a number of factors limit the risk that inaccuracies will be ongoing – for example:
- There are standard reviews – after 26 weeks or 52 weeks, depending on the benefit type – at which time the beneficiary has to re-declare their core eligibility. These reviews provide an opportunity to pick up an ongoing inaccuracy that has arisen from past provision of incorrect or incomplete information.
  - Beneficiaries receiving certain categories of benefit (such as Unemployment Benefit) are required to have client plans that involve reassessments at much more frequent intervals than the periods between the standard reviews. Any inaccuracies dating from the grant of the benefit may therefore be picked up much sooner.
  - For some (often third-tier) benefits, ongoing inaccuracy is not an issue as they may be once-only payments or they may be paid for only a short time.
- 5.48 Our examination of ARP data confirmed that many errors are small and non-recurring.
- 5.49 In 1996 the Ministry undertook a risk-sizing exercise with the aim of quantifying the amount of over-payment (including those made as the result of fraud) that may exist. The Ministry and we both consider that the method used in the 1996 exercise was flawed. However, we do see value in estimating the extent of fraud at regular intervals by an appropriate method. Such an exercise would involve taking a sample of beneficiaries and measuring the extent to which their payments match their entitlements.

### *Conclusions*

- 5.50 The ARP does not estimate the amount of under- or over-payments. Other available information and our own analysis suggest that the risk of large errors is low. However, it is important to undertake a specific exercise periodically to estimate the amount of over-payments (including fraud). The estimate would also provide the Ministry with a factual basis on which to estimate whether the current level of expenditure on Benefit Control (\$38.3 million in 2003-04) yields the greatest cost/benefit.



## ACCURACY MEASUREMENT AND REPORTING

### Recommendation 11

We recommend that –

The Ministry regularly performs a risk-sizing exercise to estimate the amount of over-payments.

## Identifying and Reporting the Causes of, and Hardship Caused by, Errors

- 5.51** When under- or over-payments were identified, the Ministry previously did not identify their causes – for example, whether they were because of an error by the beneficiary or by the case manager. In particular, inaccurate payments caused by processing errors were not clearly identified and reported. However, the Ministry has recently begun undertaking an analysis of errors, validating it against other data and communicating its findings to the regions. We applaud this initiative.
- 5.52** A systematic collection of information on errors, their size and cause, and how they were found can provide useful indicators to help management assess for policy purposes:
- priorities for focusing effort to achieve better accuracy – by identifying processing errors early and by avoiding them; and
  - the hardship that such errors might be causing beneficiaries.

### Recommendation 12

We recommend that –

The Ministry continues to explore the collection and analysis of information on errors, their size and cause, and how they were found, and to link this work with enhancements to its information technology systems.



## Part Six

# Other Opportunities for Improvement







### Introduction

**6.1** Our examination focused on the definition of accuracy as it relates to the Ministry's obligations under the Social Security Act and its Purchase Agreement. However, we identified wider opportunities for improvement outside this relatively narrow scope. In this part, we outline opportunities in two main areas:

- the consequences of the Ministry's structure in regard to accuracy; and
- obtaining more complete information of the potential benefit population.

### Consequences of the Ministry's Structure

**6.2** The Service Delivery and Specialist Services groups described in paragraph 2.6 on pages 26-27 operate in relative isolation from one another, because the Ministry views their objectives as distinct (although they do collaborate closely in a number of areas). For example:

- Debt Management, as a Specialist Services section, undertakes the recovery of debts from people who are no longer receiving any benefit. Service Delivery undertakes debt recovery from current beneficiaries.
- Benefit Control, another Specialist Services section, identifies instances of possible fraud once a benefit has been granted. Case managers undertake activities that limit the potential for fraud, but the Purchase Agreement does not require the Ministry to link the resources of Benefit Control to any assessment or estimate of the number of inaccuracies that are not identified by case managers.

**6.3** Separate funding streams for each type of activity (see Figure 12 on the next page) reinforce the distinction between objectives. Benefit Control is funded through Output Class D4, and Debt Management through Output Class D5. Service Delivery is funded through three Output Classes – D1 for working-age beneficiaries, D3 for Senior Citizen Services, and D2 for getting people into work.



## OTHER OPPORTUNITIES FOR IMPROVEMENT

*Figure 12  
Funding of Service Delivery and Specialist Services*

| SERVICE DELIVERY       |  |
|------------------------|--|
| <b>Output Class D1</b> | Income support services to working-age beneficiaries |
| <b>Output Class D2</b> | Employment services to working-age beneficiaries     |
| <b>Output Class D3</b> | Services for senior citizens                         |
| SPECIALIST SERVICES    |  |
| <b>Output Class D4</b> | Benefit control                                      |
| <b>Output Class D5</b> | Debt management – former clients                     |

### What Does the Separation Mean for Accuracy?

- 6.4 The separation of Service Delivery and Specialist Services has two consequences:
- the Ministry takes a different approach to funding and resourcing each type of activity; and
  - there is a risk that new initiatives will be developed in isolation, with insufficient consideration of whether there would be any advantages in creating complementary processes for both activities.

### Different Approaches to Funding and Resourcing

- 6.5 The funding and resourcing of Benefit Control is approached from a cost/benefit perspective. For example, Benefit Control will seek to identify over-payments (including fraud) early so that the prospective saving from doing so can be maximised. Currently, the cost/benefit target for Benefit Control is \$2.50 saved for every \$1 spent. Decisions on programmes that Benefit Control undertakes are judged against this target.





## OTHER OPPORTUNITIES FOR IMPROVEMENT

- 6.6 It would be possible for Service Delivery to adopt a similar approach. For example, if a reduction in caseloads were shown to be linked with improved accuracy (and therefore a quantifiable reduction in under- and over-payments and resulting beneficiary hardship), the Ministry could consider the cost/benefit case for extra funding to reduce caseload levels.
- 6.7 However, the Ministry does not view Service Delivery in this way, and has indicated that it has not considered putting any such case to the Government.
- 6.8 While the Ministry endeavours to keep the rate of inaccuracies to no more than 10-12%<sup>18</sup>, this target – and any incremental improvements over time – is based on what appears achievable from previous experience, rather than an informed assessment of the costs and benefits of different resourcing options.

### *Separate Development of Programmes*

- 6.9 In our view, the Ministry has not set out to devise complementary or integrated processes that achieve the best mix of the combined work of both Service Delivery and Benefit Control in meeting overall performance targets.
- 6.10 For example, Benefit Control has the primary responsibility for limiting fraud and identifying fraudulent cases. But Service Delivery activities could also be altered or expanded to help limit fraud. These could include altering target times for processing benefits<sup>19</sup> to accommodate further fraud prevention and deterrence procedures – such as more rigorous authentication.
- 6.11 However, the Ministry does not assess such options. Nor does the Ministry consider any adjustment to the level of Benefit Control activities consequent on changes in the accuracy levels actually achieved by Service Delivery. The Ministry has commented that its Purchase Agreements with the Minister have not provided for such adjustments.

<sup>18</sup> That is, to be accurate in no less than 88-90% of cases – see paragraph 5.4.

<sup>19</sup> The Ministry is currently required by the Purchase Agreement to process 95% of benefit applications within five days, although it sets higher internal targets.



## OTHER OPPORTUNITIES FOR IMPROVEMENT

### Conclusion

- 6.12 The Ministry views the Service Delivery and Specialist Services components of its benefit administration separately. Viewing them in this way runs the risk that potential cost/benefit opportunities from examining the various components together will not be identified.

#### **Recommendation 13**

We recommend that –

The Ministry treats all its processes for administering benefits as components of an integrated system. It should periodically re-estimate the optimum balance of effort between the different components in order to achieve the most cost-beneficial outcome. The method of estimation should consider all costs and benefits, including those incurred by or realised by beneficiaries, rather than being confined to those of the Ministry.

### **Obtaining More Complete Information on the Population of Potential Beneficiaries**

- 6.13 In our view, the Ministry needs information on:
- the number of people likely to be eligible for benefits but who are not claiming them; and
  - whether this is because the people are not aware of their eligibility or have chosen not to apply.
- 6.14 The Ministry requires this information if its forecasts of the number of beneficiaries and the amounts of benefit it is likely to pay out are to be as accurate as possible. The information should also be reported to Parliament.
- 6.15 The Ministry does not believe that seeking out eligible people is its responsibility in the light of its interpretation of the Ministry's obligations under the Social Security Act. However, for 2002-03, Parliament expressly appropriated money for *The provision of information on the full range of assistance available to enable people to apply for any such assistance they consider they might be eligible for and wish to apply for.*<sup>20</sup>

<sup>20</sup> *Estimates of Appropriations 2002-03, parliamentary paper B.5 Vol.II, page 1060.*





## OTHER OPPORTUNITIES FOR IMPROVEMENT

- 6.16 Making an effort to reach these eligible people would seem to be consistent with the Government's social policy goals. Indeed, the Ministry undertakes a range of activities that are intended, among a number of objectives, to provide information to people who may be eligible for a benefit but have not applied (see Figure 13 below).

*Figure 13*  
*Ministry Activities to Improve "Uptake" of Benefits*

Provision of information to potential beneficiaries, including advice:

- From case managers about possible eligibility for extra assistance e.g. Accommodation Supplement, Childcare Subsidy, and Family Support. Includes maintaining contact with people placed in work to help them achieve sustainable employment.
- On the front page of main benefit application forms about what extra assistance may be available.

Development and support of front-line staff to help them provide appropriate advice:

- Training through regional trainers, team coaches, and weekly briefings.
- On-line assistance – desk file, legislation, policy, "What's New" page outlining changes.
- Helpline – on-call advice and assistance for staff.

### Conclusion

- 6.17 These activities reinforce our view that the Ministry requires fuller information on the potential population of beneficiaries, since the activities are likely to be more effective if the characteristics of the target population are known.





## OTHER OPPORTUNITIES FOR IMPROVEMENT

### Recommendation 14

We recommend that –

The Ministry periodically undertakes exercises to estimate the number of people who are potentially eligible for social security assistance but who have not applied. For a supplementary benefit such as the Accommodation Supplement, the exercise would be likely to require information on personal income that might best be obtained from the Inland Revenue Department. The Ministry should investigate the feasibility of undertaking data matching or data extraction exercises that would yield the necessary information.



# Appendix



## APPENDIX

# Technical Issues Affecting the Accuracy Reporting Programme

## Introduction

1. The Accuracy Reporting Programme (ARP) used by the Ministry of Social Development (the Ministry) involves drawing a statistically random sample of benefits and checking each one for accuracy. Such individual checks are generally referred to by statisticians as *observations* (or sometimes as *experiments* or *trials*). For each observation in the ARP process, there are only two possible outcomes – the benefit was either *correct* or in *error*.
2. When an observed variable<sup>21</sup> can have one of two possible values, it is referred to as *binomial*. Suppose that the total number of all benefits is **N**, the true proportion of correct benefits is **P**, and the true proportion of errors is **Q**. If a sample of **n** observations are made, and an observed number **c** are correct, then the remaining observations **e** must all be errors, so that:

$$c + e = n$$

3. The collected results of a number of binomial observations can also be expressed as proportions, so that:

$$c/n = p$$

$$e/n = q$$

$$p + q = 1$$

4. There are at least three significant reasons why the observed proportions **p** and **q** might not be the true proportions **P** and **Q**. These are:

- sampling error;
- measurement reliability; and
- non-sampling errors.

In this appendix we will provide some explanation of these possible sources of error and how two of them can be quantified.

21 A common form of notation used in statistical literature is that a population parameter is denoted by an unembellished variable such as "**p**" and an estimate of that parameter obtained from a sample of the population is denoted by adding a circumflex – "**p̂**". We have not used the common notation in this explanation in the belief that readers unfamiliar with mathematical notation often find it confusing until they have had time to become familiar with it.





## Sampling Error

5. Since its value is derived from random sampling, it is not likely, except by lucky accident, that the observed proportion  $p$  of correct benefits in the sample will exactly equal the true proportion  $P$  of correct benefits in the total “population” of all benefits (and likewise, that the sample proportion  $q$  will exactly equal the true proportion  $Q$ ).
6. For example, we might choose a sample  $n$  of 1000 benefits from a total population  $N$  of 1,000,000 benefits. The true proportion  $P$  of correct benefits in the total population might be 0.9 (that is, 90%) and the true proportion of errors ( $Q$ ) might be 0.10 (10%). However, because it is chosen at random, our particular sample of 1000 benefits might disclose a sample proportion  $p$  of correct benefits of 0.87 (87%) and a sample proportion  $q$  of errors of 0.13 (13%).
7. Given that the proportions observed from a sample are not likely to be exactly the same as the true proportions, it would be very useful to have some idea of how far the observed values are likely to differ from the true values.
8. Intuitively, we might realise that if the sample size  $n$  is very large in relation to the population size  $N$ , the observed values of  $p$  and  $q$  are likely to be very close to  $P$  and  $Q$  (because, if the sample size  $n$  was actually equal to  $N$  – and assuming that our observations were always accurate – by definition,  $p$  would exactly equal  $P$  and  $q$  would exactly equal  $Q$ ). Similarly, if  $n$  is very small in relation to  $N$  then the chances of a significant difference between  $p$  and  $P$  (and between  $q$  and  $Q$ ) are much higher. The bigger the sample, the more likely the estimate is to be accurate. But how much more likely?
9. Statisticians have devised a useful method for assessing how likely it is that  $p$  is a good estimate of  $P$ . The method is known as a *confidence interval*.
10. A confidence interval is a span of possible values of  $P$ . It can be calculated by a formula that has been devised to ensure that, in a given high percentage of cases, the interval will include the true value  $P$ . In the example already cited, if an observed value is 0.87, some confidence interval might be a span of possible values of  $P$  that ranged from 0.82 to 0.92 (which, in this example, does indeed contain the true value of  $P$ , namely 0.90).





## APPENDIX

11. To better explain the concept of a confidence interval, we need to develop the previous example further. From our particular sample of 1000 benefits, we obtained observed proportions  $p = 0.87$  and  $q = 0.13$ . If we had drawn some other random sample of the same size, we might have obtained a slightly different pair of observed values – say,  $p = 0.89$  and  $q = 0.11$ . If we drew a third sample of the same size, we might have obtained yet another set of observed values – say,  $p = 0.91$  and  $q = 0.09$ . (Remember that the true values in this example are  $P = 0.90$  and  $Q = 0.10$ .) If we took numerous samples of 1000 benefits, we would expect to find our observed values of  $p$  and  $q$  clustered around the true values  $P$  and  $Q$ . If we took very many such samples, we would expect to find most observed values of  $p$  and  $q$  were quite close to  $P$  and  $Q$ , and only a few observed values of  $p$  and  $q$  which were significantly different from  $P$  and  $Q$  (because in those few cases we happened to draw a quirky sample).

12. Because:

- each observed value of  $p$  is particular to the actual sample from which it was observed; and
- each observed value of  $p$  will in general be slightly different from the values of  $p$  observed from other samples;

the confidence interval calculated from a particular sample will in general be slightly different from the intervals calculated in the same way but using the observed values from other samples. However, we can decide in advance how sure we want to be that the span of our confidence interval will be wide enough to contain the true value  $P$ . We can then calculate a confidence interval of the appropriate width.

13. For example, if we specify a 95% confidence interval, we can calculate a span of possible values of  $p$  that, in 95% of cases, will include the true value of  $P$ . In other words, given a total population of  $N$ , if we had drawn all the very many possible samples of size  $n$  and calculated confidence intervals for all the observed values of  $p$ , 95% of those confidence intervals would contain the true value  $P$ .
14. There is no particular reason why we should have chosen 95% (although that is a commonly-used figure). We could have chosen some other percentage – for example, 90% if we didn't need to be quite so sure that the interval contained the true value, or 99% if we needed to be very sure that the interval contained the true value.





## APPENDIX

15. A number of methods have been devised for calculating the confidence interval of a proportion and there has been considerable technical debate about the best method. Computer simulations have shown that all methods have some flaws, including the so-called “exact” methods. (These have also been shown not to be exact at all, but conservative approximations.) The precise extent of the discrepancies associated with each method varies, depending on the values of  $p$  and  $n$ . However, if the sample size is reasonably large, most methods will produce very similar confidence intervals. The difficulties come when the sample size is small and the observed proportions are very low or very high.
16. The most commonly used method for calculating a 95% confidence interval for an observed proportion was devised by Abraham Wald. It applies the following formula:
 
$$\text{Lower Limit} = p - 1.96 \sqrt{\frac{pq}{n}}$$

$$\text{Upper Limit} = p + 1.96 \sqrt{\frac{pq}{n}}$$
17. However, there are some difficulties with the Wald method. For example, it does not work at all without modification in cases where:
  - the formula produces a limit for the confidence interval that is less than 0 or greater than 1 (implying a negative proportion or a proportion greater than 100%, both impossible); or
  - the observed proportion is zero.
18. Most textbooks on statistics teach the Wald method and assert that it is an acceptable method if the sample size  $n$  is reasonably large and  $p$  is not too close either to 0 or 1. However, recent research (for example, Brown, Cai and Das Gupta, 2001) has shown it to be much more erratic across wider ranges of  $p$  and  $n$  than had previously been understood.
19. Agresti and Coull (1998) devised a method they term the “modified Wald” method. It is computationally similar to the Wald method but uses simple adjustments to the values of  $p$ ,  $q$  and  $n$ . Computer simulations have shown it to be more accurate in most circumstances than other methods, including the so-called “exact” methods. It is the method recommended by Brown, Cai and DasGupta for cases where  $n \geq 30$  and accordingly we have adopted it for use in this study.



## APPENDIX



20. The necessary modifications for a 95% confidence interval are:

$$p' = \frac{c+2}{n+4}$$

$$q' = \frac{c+2}{n+4}$$

$$\text{Lower Limit} = p' - 2\sqrt{\frac{p'q'}{n+4}}$$

$$\text{Upper Limit} = p' + 2\sqrt{\frac{p'q'}{n+4}}$$

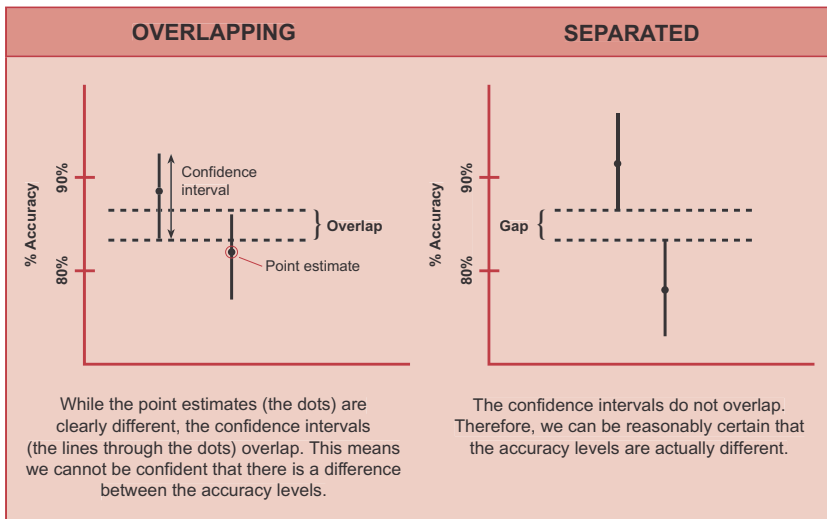
21. If the lower limit calculated using that equation is less than 0, set the lower limit to 0. Similarly, if the upper limit is greater than 1, set the upper limit to 1.
22. For any values of  $p$  and  $n$ , at least 92% – and generally around 95% – of intervals calculated by using this formula will contain the true result. It also deals more effectively with cases where the observed proportion  $p$  is 0.
23. It is possible to use confidence intervals to make judgements about whether or not it is safe to assume that the error proportions in two separate populations of benefits are actually different. For example, if the ARP-determined level of benefit accuracy in any given year is 91% and in the next year is 90%, are we safe in concluding that the level of benefit accuracy has actually declined from the previous year?
24. The simple rule is that if the confidence intervals do not overlap, it is safe at the given level of confidence to conclude that the true proportions are indeed different. However, if two confidence intervals overlap, the difference between the observed proportions might or might not be statistically significant – without further scrutiny it is unsafe at that level of confidence to conclude that the true proportions are actually different.<sup>22</sup>

<sup>22</sup> Another approach is to calculate a confidence interval for the difference between the two proportions rather than calculating confidence intervals for each proportion separately.





## APPENDIX



25. To return to the example just cited, the 95% confidence interval for the national ARP measure of benefit accuracy is approximately  $\pm 1\%$ . The 95% confidence interval for the first year is therefore 90% – 92% and for the second year is 89% – 91%. Since these confidence intervals overlap (across the range 90% – 91%), it is unsafe to conclude at the 95% level of confidence that accuracy levels have in fact declined. However, if the ARP measure for the second year was 88%, (and the confidence interval was therefore 87% – 89%), it would be safe at the 95% level of confidence to conclude that accuracy levels had actually declined.

## Measurement Reliability

26. In measurement theory, there are several ways of defining and interpreting the concept of the *reliability* of a method of measurement<sup>23</sup>. For most practical purposes, however, the concept is motivated by a simple question – If we measure a particular item again using the same method of measurement, how likely are we to get the same result?

<sup>23</sup> A more detailed treatment is beyond the scope of this appendix. More information can be obtained from any text book on measurement theory – for example, Allen, M. J. and Yen, W. M., *Introduction to Measurement Theory*, Brooks Cole, 1979.



## APPENDIX



27. For example, repeatedly measuring a particular length of wood with a well-calibrated wooden ruler might produce highly consistent results. In such situations, the measurement is said to be *reliable*. However, repeatedly measuring the same length of wood with a tape measure made from elastic might produce highly inconsistent results, because small changes in tension may cause significant stretching or contracting of the tape. In such situations, the measure is not reliable. Most methods of measurement are unreliable to some extent, and it is useful to have some idea of how reliable or unreliable they actually are.
28. In relation to the ARP, the most likely reason that the observed value  $p$  will differ from the true value  $P$  (apart from sampling error) is *measurement error*. For each observation made (i.e. each benefit sampled), an assessor makes a judgement about whether or not the benefit is correct. But the assessor's judgement may itself be wrong. In reporting ARP measurements of benefit accuracy, it would be useful to present some idea of how likely it is that some of the assessors' judgements were wrong.
29. It seems to us that there are two possible approaches.
  - estimating and reporting the standard error of measurement; and
  - estimating and reporting inter-assessor reliability.
30. The concept of a *standard error of measurement* is derived from a set of assumptions and consequential deductions that are known as "true score theory" (TST). Among other assumptions, TST assumes that a measurement has a "true" score and that actual measurement errors are normally distributed<sup>24</sup> around the true score. The span of possible values for the true score that is one standard deviation wide is known as the standard error of measurement.
31. It is rare that a large number of repeated measurements of a particular value have been made, so that the standard deviation of the measurement error distribution can be estimated directly. However, by making the further assumption that the measurement error distributions of all measurements in a particular sample are the same, it is possible to estimate the standard deviation of that common distribution from the observed correlation between single repeat measurements of each observation in the sample.

24 That is, distributed in accordance with the bell-shaped "normal" frequency distribution.





## APPENDIX

32. The standard error of measurement is a commonly-used metric of reliability, but it is sensitive to a number of assumptions. On balance, we don't favour its use in ARP reporting.
33. A second approach is to measure the *inter-assessor reliability*<sup>25</sup>. This can be done by having a number of assessments rechecked by an independent second assessor (who should be unaware of the original assessment). The inter-assessor reliability can then be determined by calculating the correlation between assessors' judgements. A highly correlated set of judgements implies a high degree of reliability in the observed levels, while a weakly correlated set of judgements implies that the observed levels of benefit accuracy will be somewhat uncertain.
34. We consider that reporting inter-assessor reliability would constitute a useful enhancement of the ARP, and we recommend its introduction.

### Non-Sampling Errors<sup>26</sup>

35. ARP accuracy figures, even at national level, make no allowance for non-sampling errors, measurement errors, or any other possible types of errors that may be related to misapplied technical methods, for example. We are unaware of any study or report by the Ministry on such matters. However almost no survey can escape at least some non-sampling errors. We therefore recommended that a study be undertaken by the Ministry to assess the extent of such problems, and a report prepared.
36. Non-sampling errors can take a very wide variety of forms. See for example Lesser et. al., 1992. Finding non-sampling errors, assessing their importance, and dealing with them in detail, would require considerable interaction with staff of the Ministry, and is consequently beyond the scope of this report.

---

<sup>25</sup> Also commonly known as "inter-rater agreement".

<sup>26</sup> The remainder of the analysis and associated conclusions in this appendix were provided by our technical referee, Professor Stephen Haslett.





## APPENDIX

### Technical Issues Affecting the 5+5 Checks of Case Managers' Performance

#### Accuracy of the 5+5 Checks

37. The 5+5 procedure itself, as applied once only to individual staff, uses sample sizes that are too small to provide a satisfactory measure of individual staff performance. To illustrate, consider a staff member who meets a 90% accuracy (i.e. 10% error) requirement in terms of errors made in cases/files for which s/he is responsible. Then (ignoring the equivalent of finite populations corrections for simplicity, and under simple random sampling) the probabilities that this staff member has 0, 1, 2, 3, 4, 5 errors found in five sampled files are given by:

| Number of errors<br>in sample of<br>size 5+5=10 |                               | Probability given<br>real error rate<br>meets 10% standard |
|---|-------------------------------|--|
| 0   | $(0.9)^5 =$                   | 0.59049  |
| 1   | $5*(0.1)*(0.9)^4 =$           | 0.32805  |
| 2   | $((5*4)/2)*(0.1)^2*(0.9)^3 =$ | 0.07290  |
| 3   | $((5*4)/2)*(0.1)^3*(0.9)^2 =$ | 0.00810  |
| 4   | $5*(0.1)^4*(0.9) =$           | 0.00045  |
| 5   | $(0.1)^5 =$                   | 0.00001  |

38. Hence, the probability of finding two or more errors in a sample of size five, when the staff member actually meets a 90% error-free standard, is nearly 0.1 (which would penalise one such staff member in every 10), and the probability that the sample error rate is higher than the actual error rate is more than 0.4. This analysis ignores any incorrect assessments of whether a file is 'in error', but this fact does not allow the central problem (detailed below) to be avoided.





## APPENDIX

39. The complication is that if a staff member's actual error rate was unacceptable, say only 75% of his/her files were accurate, then the probability that this remains undetected by a sample of five of his/her cases can be derived as follows:

| Number of errors<br>in sample of<br>size 5+5=10 |                                 | Probability given<br>real error rate<br>meets 25% standard |
|---|---------------------------------|--|
| 0   | $(0.75)^5 =$                    | 0.23730  |
| 1   | $5*(0.25)*(0.75)^4 =$           | 0.39551  |
| 2   | $((5*4)/2)*(0.25)^2*(0.75)^3 =$ | 0.26367  |
| 3   | $((5*4)/2)*(0.25)^3*(0.75)^2 =$ | 0.08789  |
| 4   | $5*(0.25)^4*(0.75) =$           | 0.01465  |
| 5   | $(0.25)^5 =$                    | 0.00098  |

40. Hence, if zero or one error in the sample is acceptable (as from the previous analysis it certainly needs to be for good staff not to be incorrectly penalised more than one time in 10), then a staff member with an unacceptable error rate of one error in every four files has a considerably better than even chance (i.e. a probability of  $0.23730+0.39551=0.63281$  of going completely undetected).
41. While the situation is rather better if all 5+5=10 files are considered together (see below), there remains a strong possibility that, when using a standard that does not penalise good staff too frequently, the errors of others go substantially undetected. This is clearly undesirable.
42. In principle, an improved procedure would be to use some form of sequential sampling, where, if an unacceptable sample error level is found for a given staff member, a further sample is taken. This procedure would certainly require specialist advice to design and analyse, however, because the usual binomial or other tabulated probabilities (such as those above) are not conditional probabilities which are required at any additional phase of sampling, and so specialised calculations (and software programs) would be needed for implementation.
43. However, it is very clear from this analysis that the present 5+5 scheme needs modification and larger sample sizes, if it is not to be rather arbitrary in its assessment of individual staff standards, at least when applied for a single period.





## APPENDIX



44. If the 5+5 samples are treated as a single sample of size 10, it is instructive to consider the probability of finding a particular number of errors in the sample for a particular case manager, under two different scenarios:
1. given the case manager's work actually meets a 10% error standard,
  2. given the case manager's work really has 25% of benefit files in error.

| Number of errors in sample of size 5+5=10 |  | Probability when actual error rate meets 10% standard |
|---|--|---|
| 0   | $(0.9)^{10} =$                         | 0.34868   |
| 1   | $10*(0.1)*(0.9)^9 =$                   | 0.38742   |
| 2   | $((10*9)/2)*(0.1)^2*(0.9)^8 =$         | 0.19371   |
| 3   | $((10*9*8)/(3*2))*(0.1)^3*(0.9)^7 =$   | 0.05740   |
| 4   | $(10*9*8*7)/(4*3*2)*(0.1)^4*(0.9)^6 =$ | 0.01116   |

| Number of errors in sample of size 5+5=10 |  | Probability given real error rate is 25% |
|---|--|--|
| 0   | $(0.75)^{10} =$                          | 0.05631                                  |
| 1   | $10*(0.25)*(0.75)^9 =$                   | 0.18771                                  |
| 2   | $((10*9)/2)*(0.25)^2*(0.75)^8 =$         | 0.28157                                  |
| 3   | $((10*9*8)/(3*2))*(0.25)^3*(0.75)^7 =$   | 0.25028                                  |
| 4   | $(10*9*8*7)/(4*3*2)*(0.25)^4*(0.75)^6 =$ | 0.14600                                  |

45. Hence, if three or fewer errors in the sample of 10 is used as the criterion for acceptability, for example, then the chance of misjudging a staff member who actually meets a 10% error rate standard is about 0.07, i.e. about 1 in 14 such staff members would be wrongly judged as having unacceptable error rates. At this same standard however, a staff member whose error rate was 25% (i.e. 2.5 times higher than the 10% limit) would have a probability of 0.526 (i.e. a better than even chance) of not being detected (i.e. less than half the staff with a 25% error rate will be detected).





## APPENDIX


46. The statistical conclusion is that a single 5+5 sample is too small to provide a fair assessment of case managers' benefit file error rates, even if all 10 benefit files are combined for each case manager rather than treated as two separate groups of five. The problem is that the 'statistical power' of the test is low because sample sizes are too small.
47. A situation which is not as imprecise is when a case manager's 5+5 assessments are aggregated over a period. Because the period, and hence the number of assessments, may vary, further detailed analysis that meets all circumstances is not possible. However, the general methodology above will yield the numerical answers when it is used in combination with the relevant data.

### Issues Associated with Combining ARP Data with Data from the 5+5 Checks of Case Managers' Performance

48. How can ARP and 5+5 sampling systems best be linked? One method involves what is called 'double sampling'. A fundamental reference is Cochran (1977), especially Chapter 12, pages 327-358. Extension of these methods to the sampling schemes used for ARP and 5+5 sampling (at least within regional offices) would be necessary before implementation.
49. How double sampling works can be illustrated with a simple biological example. Consider the problem of accurately measuring the surface area of all the leaves on a deciduous tree. Measuring the surface area of every leaf is not feasible. However, if a comparatively small number of leaves are taken, the surface area and dry weight determined for each, and the (regression) relationship between surface area and dry weight is estimated, all that is necessary after finding this relationship is to wait for all the leaves to fall, dry and weigh them. The regression relationship can then be used to estimate surface area from weight. This method is comparatively inexpensive.
50. A parallel double sampling scheme for assessing error rates in the Ministry's benefit files would be to use the relationship between 5+5 results and ARP results applied to a comparatively small sample of the *same* files, as the basis for combining all 5+5 results with the ARP results, to get more accurate benefit file error rate assessments. A sub-sample assessed by both methods would be used for calibration.



## APPENDIX

- 
51. If the Ministry was to adopt double sampling, further external technical input is highly recommended, particularly since the current sampling schemes for 5+5 and ARP differ. Detailing the required survey design for double sampling is beyond the scope of this report. However, the method looks promising and we recommend that it be investigated further.

## Sampling Issues

52. In this report, confidence intervals have been calculated on the assumption that sampling, including sampling within regions, is conducted in such a manner as will achieve a “simple random sample” (srs).
53. It is not completely clear whether sampling within regions is on an srs basis. If sampling within regions is not by srs, it will have some effect on confidence intervals, although the effect should be slight for any common, easily implemented sampling scheme in the absence of marked sample clustering and any non-sampling problems.
54. ARP sampling on a national basis uses stratified sampling by regional offices with proportional allocation for each office, based on the distribution of applications and reviews for the financial year 1995-96. The Ministry has stated that the variation in Applications and Reviews is not significant over time. This sampling scheme is not quite equivalent to an srs, although differences in estimated size of a confidence interval should be slight, so that the srs formula for a national confidence interval can be used as an approximation (given similarity of assessments of benefit assessment accuracy across offices) provided sampling within offices is by srs.

## References

- Agresti, A. and Coull, B. (1998). *Approximate is Better than “Exact” for Interval Estimation of Binomial Proportions*, **The American Statistician**, 52: 119–126.
- Allen, M.J. and Yen, W. M., *Introduction to Measurement Theory*, Brooks Cole, 1979.
- Brown, L.D., Cai, T.T. and Das Gupta, A., *Interval Estimation for a Binomial Proportion*, **Statistical Science**, 16, 2: 101–133.
- Cochran, W.G., *Sampling Techniques*, 3<sup>rd</sup> Edition, John Wiley and Sons, 1977.
- Haslett, S and Telfar, G., *Sample Designs for Auditing Error Data: A Case Study*, **New Zealand Statistician**, June 1995, Vol. 30, No. 1, pp 23–36.
- Lesser, J.T. and Kalsbeek, W.D. *Nonsampling Error in Surveys*, John Wiley and Sons, 1992.



## Recent Publications by the Auditor-General

Other publications issued by the Auditor-General in the past 12 months have been:

- Inquiry into Public Funding of Organisations Associated with Donna Awatere Huata MP
- Managing Threats to Domestic Security
- Annual Report 2002-03 – B.28
- Co-ordination and Collaboration in the Criminal Justice Sector
- Local Government: Results of the 2001-02 Audits – B.29[03b]
- Inland Revenue Department: Performance of Taxpayer Audit
- Management of Hospital-acquired Infection
- Central Government: Results of the 2001-02 Audits – B.29[03a]
- Disposal of 17 Kelly Street by Institute of Environmental Science and Research Limited
- ACT Parliamentary Party Wellington Out-of-Parliament Offices
- Annual Plan 2003-04 – B.28AP(03)
- New Zealand Defence Force: Deployment to East Timor – *Performance of the Health Support Services*
- New Zealand Defence Force: Deployment to East Timor – *Performance of the Helicopter Detachment*
- Department of Conservation: Administration of the Conservation Services Programme
- Certain Matters Arising from Allegations of Impropriety at Transend Worldwide Limited
- Management of Biosecurity Risks: Case Studies
- Ministry of Agriculture and Forestry: Management of Biosecurity Risks
- *All about ... The Controller and Auditor-General*

### Web Site

All these reports are available in PDF form on our web site [www.oag.govt.nz](http://www.oag.govt.nz). They can also be obtained in hard copy on request – [reports@oag.govt.nz](mailto:reports@oag.govt.nz). A cost may apply for hard copies.

### Subscription for Notification of New Reports

We offer a subscription facility for people to be notified by e-mail when new Reports and Latest News are added to the web site. The link to this subscription service is on our Home Page and also in the Reports section of the web site.



# **Social Security Benefits: Accuracy of Benefit Administration**

---

**Controller and Auditor-General**  
*Tumuaki o te Mana Arotake*

**ISBN 0-478-18104-3**